(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENT- UND MARKENAMT

(12) **Offenlegungsschrift**

(10) **DE 198 46 493 A 1**

(21) Aktenzeichen: 198 46 493.2
(22) Anmeldetag: 9. 10. 1998
(43) Offenlegungstag: 13. 4. 2000

(51) Int. Cl.[7]:
C 12 N 15/52
C 12 N 15/63
C 12 N 1/21
C 07 D 493/04

DE 198 46 493 A 1

(71) Anmelder:
Gesellschaft für Biotechnologische Forschung mbH (GBF), 38124 Braunschweig, DE

(74) Vertreter:
Patentanwälte Dr. Boeters, Bauer, Dr. Forstmeyer, 81541 München

(72) Erfinder:
Beyer, Stefan, Dr., 38124 Braunschweig, DE; Müller, Rolf-Joachim, Dr., 38124 Braunschweig, DE

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

(54) DNA-Sequenzen für die enzymatische Synthese von Polyketid- oder Heteropolyketidverbindungen

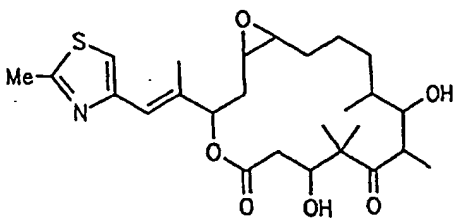
(57) Die Erfindung betrifft eine DNA-Sequenz, deren Expressionsprodukte eine enzymatische Biosynthese, Mutasynthese oder Partialsynthese von Polyketid- oder Heteropolyketitverbindungen bewirken oder daran beteiligt sind.

DE 198 46 493 A 1

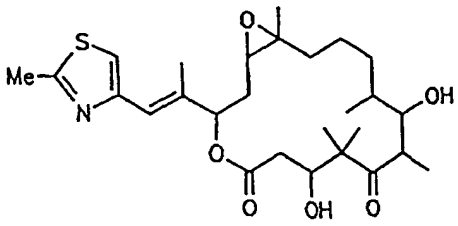
Beschreibung

Die Erfindung betrifft DNA-Sequenzen für die enzymatische Synthese von Polyketid- oder Heteropolyketidverbindungen nach Patentanspruch 1, und zwar insbesondere zur enzymatischen Synthese von Epothilonen.

Polyketid- oder Heteropolyketidverbindungen, insbesondere Epothilone, der folgenden allgemeinen Strukturformel sind beispielsweise aus DE 41 38 042, DE 196 47 580.5 und DE 197 07 501.6 bekannt:



EPOTHILON A



EPOTHILON B

worin R^1 Wasserstoff, C_{1-4}-Alkyl, C_{1-4}-Acyl, Li^+, K^+, Na^+, $^1/_2\ Mg^{2+}$ oder $^1/_2\ Ca^{2+}$ bedeutet und R^2 Wasserstoff oder eine Methylgruppe darstellt.

Die Epothilone werden in die Typen A bis F eingeteilt. Sie haben cytotoxische und/oder immunsupressive sowie antibiotische und antifungale Wirkungen und finden daher zum Beispiel Anwendung als Mittel für den Pflanzenschutz in Landwirtschaft, Forstwirtschaft und/oder im Gartenbau.

Die Epothilone wurden bisher fermentativ durch Kultivierung von Sporangium-Stämmen hergestellt und durch Anwendung herkömmlicher Techniken isoliert und gereinigt, vgl. z. B. DE 41 38 042.8.

Fermentative Techniken sind aber oft mit Nachteilen verbunden. Der produzierende Mikroorganismus erlaubt nicht in jedem Fall die fermentative Herstellung in großem Maßstab. Häufig kommt es zu Komplikationen bei der großmaßstäblichen Kultivierung oder die Ausbeuten sind gering oder die Isolierung und Reinigung sind aufwendig.

Daher wäre es vorteilhaft, wenn zur fermentativen Herstellung der gewünschten Verbindungen ein gut charakterisierter und leicht zu handhabender Mikroorganismus zur Verfügung stünde. Wenn ein solcher aber nicht in der Natur gefunden oder gezüchtet werden kann, bleibt nur noch die entsprechende Veränderung eines geeigneten Mikroorganismus mit gentechnischen Methoden. Dazu ist aber die Isolierung und Charakterisierung der entsprechenden Gene erforderlich.

Aufgabe der Erfindung ist daher gemäß Patentanspruch 1 die Bereitstellung einer DNA-Sequenz, deren Expressionsprodukte die enzymatische Biosynthese, Mutasynthese oder Partialsynthese von Polyketid- oder Heteropolyketidverbindungen bewirken oder daran beteiligt sind.

Durch die Bereitstellung einer derartigen DNA-Sequenz lassen sich folgende Vorteile erzielen.

Die DNA-Sequenz läßt sich mit üblichen molekularbiologischen Methoden in bekannte und optimierte Expressionsvektoren insertieren, wodurch die entsprechende Transformation, Selektion und Klonierung von Zellen möglich ist, die dann zur Synthese von Polyketid- oder Heteropolyketidverbindungen durch Fermentation in der Lage sind. Wenn ein überproduzierender Zellklon gewählt wird, lassen sich die gewünschten Polyketid- oder Heteropolyketidverbindungen leicht in großen Mengen herstellen und gewinnen.

Die Kenntnis der Lage der regulatorischen DNA-Abschnitte und der einzelnen Strukturgene gestattet die gezielte Mutagenese ("site-directed mutagenesis") mit üblichen gentechnischen Methoden und somit die Konstruktion von optimierten Enzymen ("protein engineering") zur fermentativen Synthese von Polyketid- oder Heteropolyketidverbindungen.

Die Erfindung betrifft somit ferner einen rekombinierten Expressionsvektor nach Patentanspruch 8, damit transformierte Zellen nach Patentanspruch 9 sowie ein Verfahren zur enzymatischen Biosynthese, Mutasynthese oder Partialsynthese von Polyketid- oder Heteropolyketidverbindungen nach Patentanspruch 15.

Vorteilhafte Ausführungsformen der Erfindung sind Gegenstand der Unteransprüche.

Die Erfindung wird nachstehend detaillierter erläutert.

Fig. 1 ist eine Restriktionskarte der erfindungsgemäßen DNA-Sequenz, die auch die Lage der regulatorischen DNA-Abschnitte und der einzelnen Strukturgene ("open reading frames" (ORF) 1 bis 14) angibt.

Fig. 2 ordnet den ORF 1 bis 14 die jeweilige biologischen Funktion (Regulatoren, Enzyme) zu.

Isolation und Charakterisierung der DNA-Sequenz

Es wurde genomische DNA aus dem Myxobakterium Sorangium cellulosum Soce90, Stamm und Anzucht bekannt aus DE 41 38 042, verwendet.

Genomische DNA wurde mit Hilfe des Qiagen Blood & Cell Culture DNA Kits (Qiagen, Hilden, FRG) isoliert. Siehe dazu "Genomic DNA handbook" S. 31 ff (Qiagen 1995). Modifizierungen: Nach Denaturierung und Proteolyse wurde eine Phenol-Chloroformextraktion, gefolgt von einer Ethanolpräzipitation, vorgenommen (Sambrook J., Fritsch E. F., Maniatis T., Molecular Cloning, A Laboratory Manual, Cold Spring Harbour Laboratory Press, New York; 1989). Die in Puffer gelöste DNA wurde anschließend auf die Qiagen-Reinigungssäulen aufgetragen und gemäß den Herstellerangaben isoliert.

Konstruktion von geeigneten rekombinierten Expressionsvektoren

Expression in Myxobakterien

Eine heterologe Expression der in **Fig.** 1 aufgeführten ORFs wird unter Verwendung eines Derivats des Plasmids pSUP102 (Simon R., Priefer U., Pühler A.; Methods in Enzymol, 118: 643–659; 1986), bei welchem das Chloramphenicolresistenzgen durch eine Streptomycinresistenzgenkassette mit Promotorelement aus dem Transposon TNS ausgetauscht wurde, ausgeführt. Homologe kurze Abschnitte genomischer DNA aus dem Wirtsorganismus werden mit den DNA-Sequenzen entsprechend **Fig.** 1 unter Benutzung effektiver Regulationselemente in z. B. die Restriktionsschnittstelle EcoRI des Vektors ligiert. Nach Amplifikation des Vektors in Escherichia coli erfolgt der Transfer der. DNA durch Elektroporation der Wirtszellen oder durch Konjugation mit Escherichia coli S17-I (Simon R., Priefer U., Pühler A., Bio/Technology 1: 784–791; 1983).

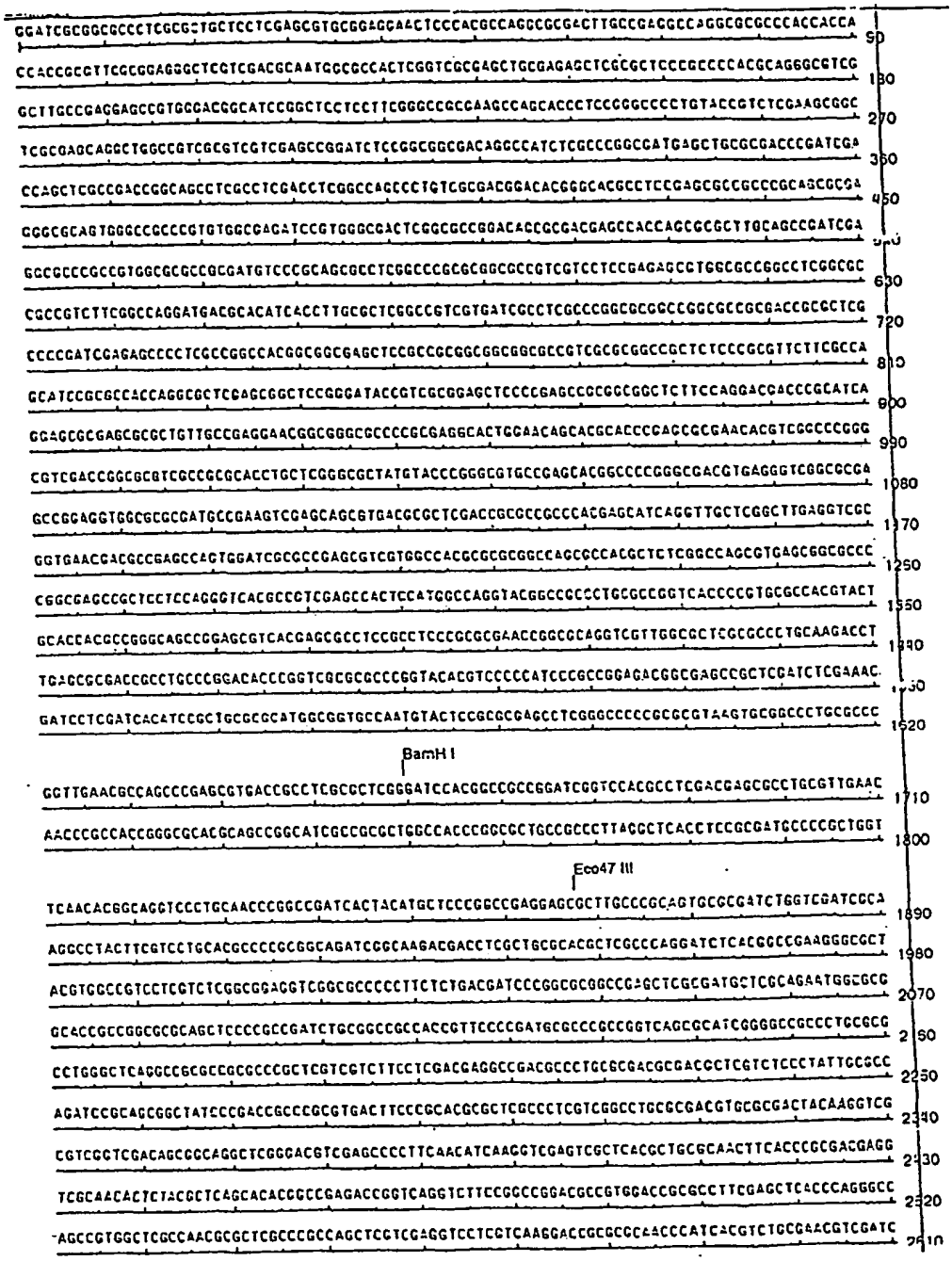
Mit Hilfe der durch den Vektor vermittelten Tetrazyklin- bzw. Streptomycinresistenz werden die Wirtszellen auf Integration der rekombinanten Plasmid-DNA durch homologe Rekombination in das Chromosom überprüft.

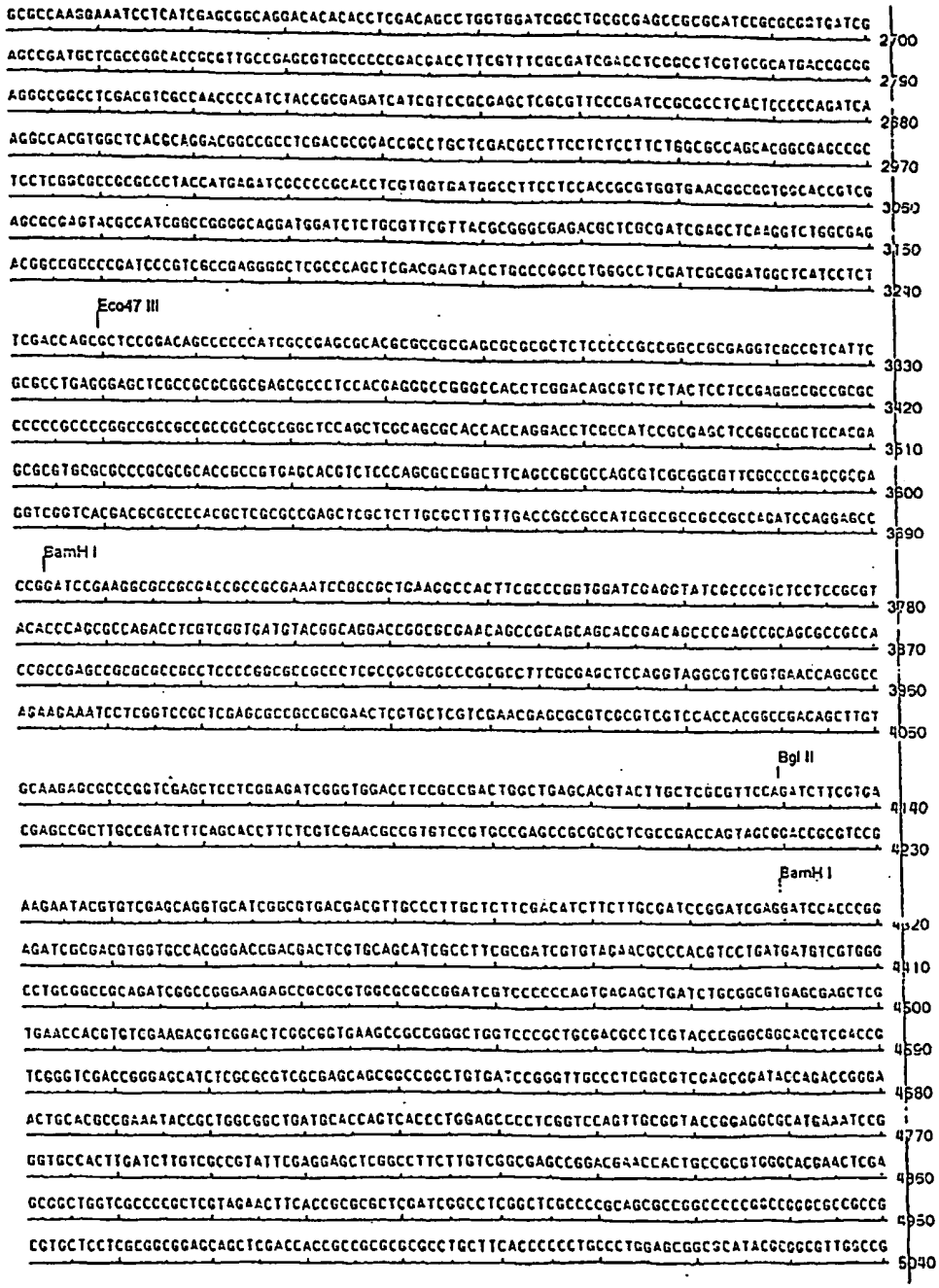
Expression in Zellen von Streptomyces

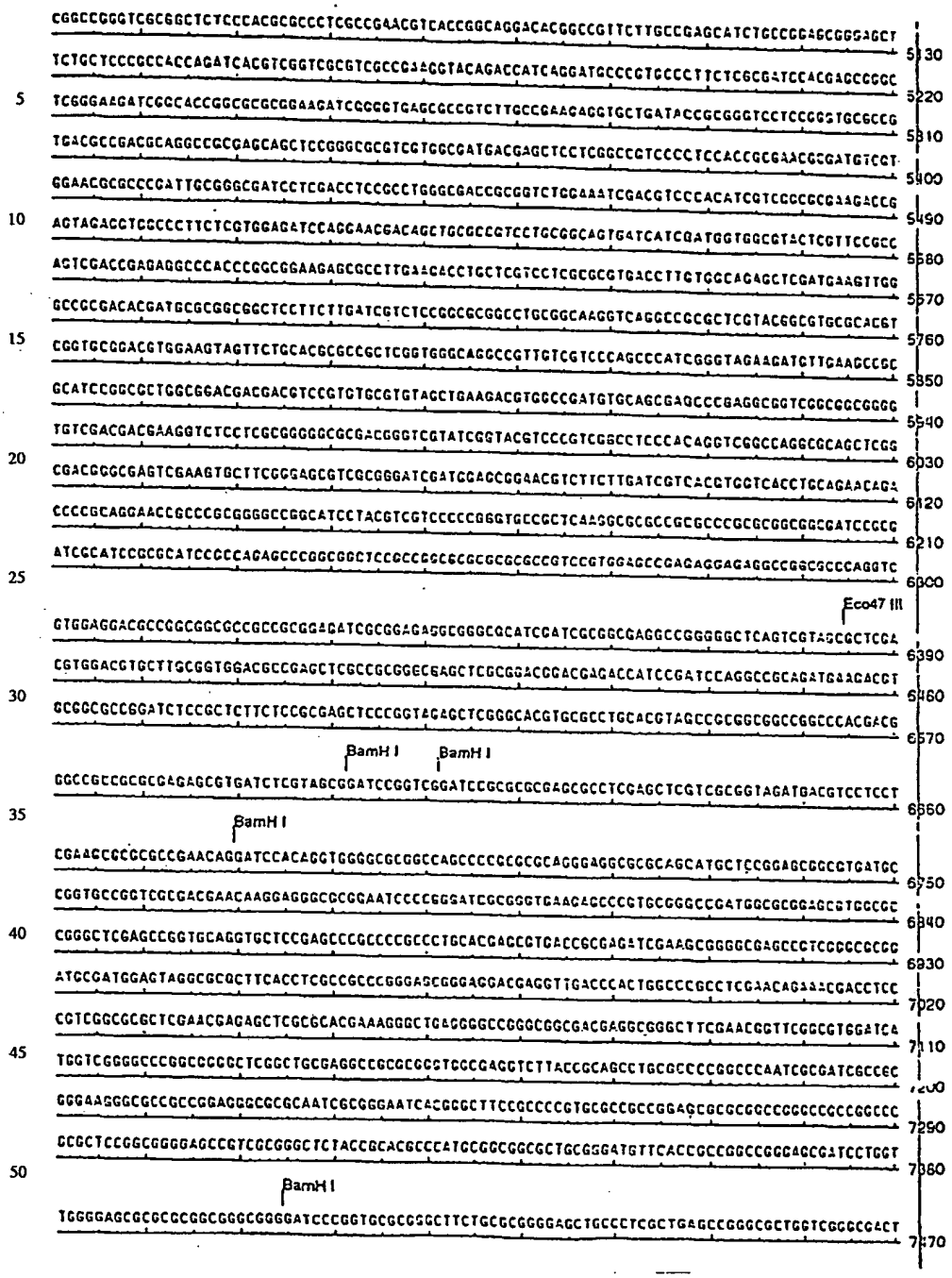
Eine heterologe Expression der in **Fig.** 1 aufgeführten ORFs wird unter Verwendung der bifunktionalen Streptomyces-Escherichia coli-Cosmide pKU206 oder pOJ466 vorgenommen.

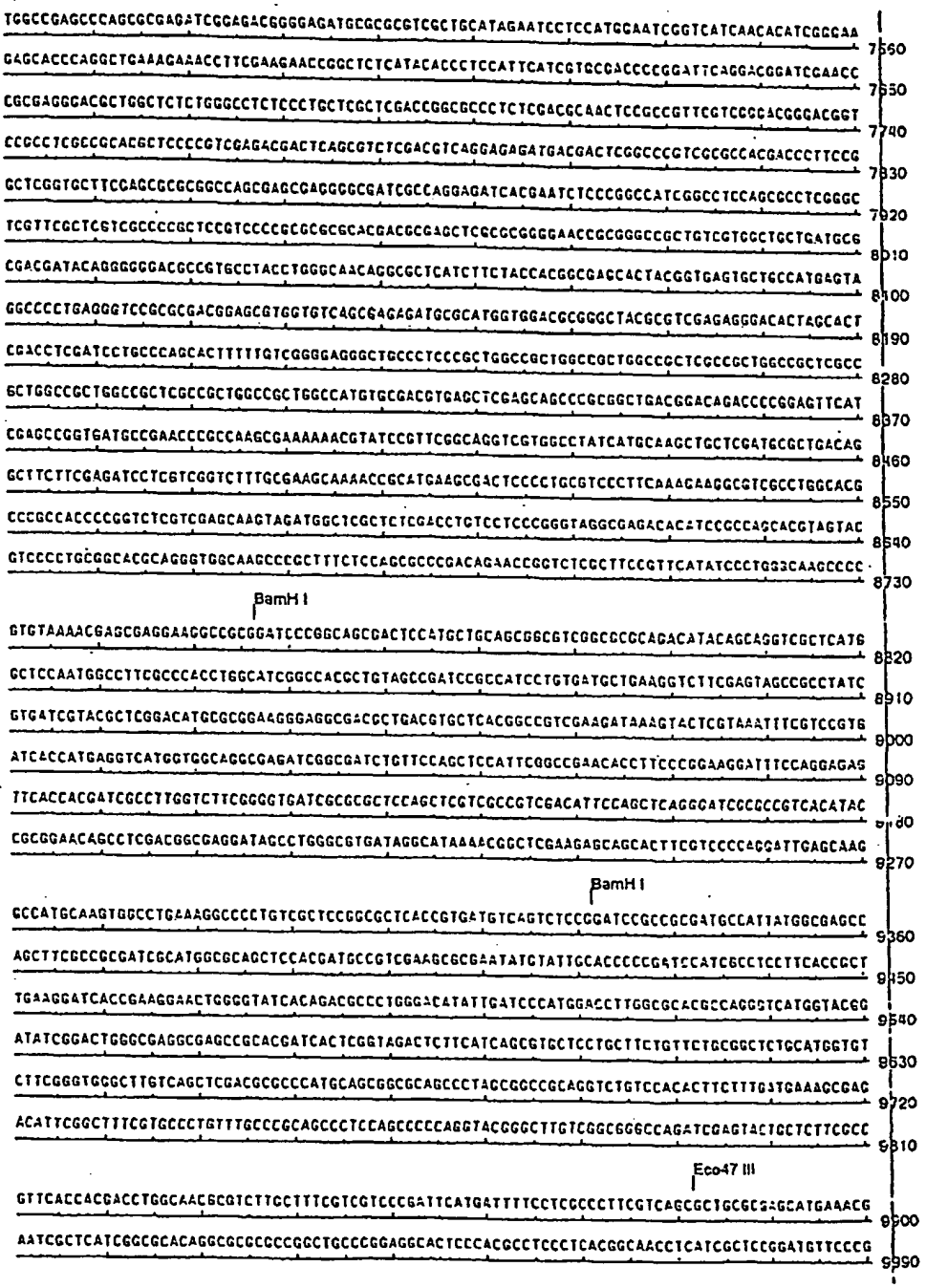
Expression von Zellen in Escherichia coli

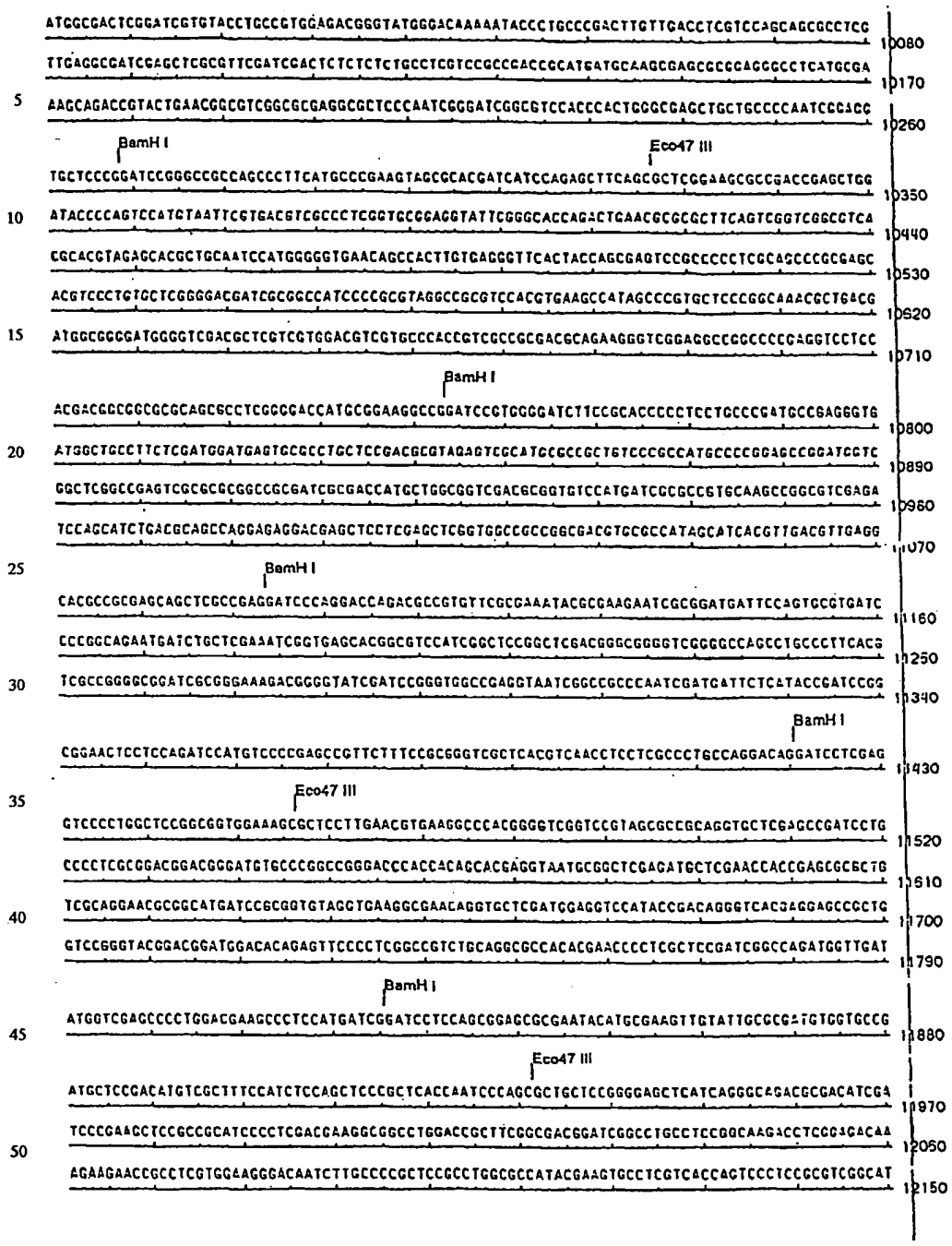
Eine heterologe Expression der in **Fig.** 1 aufgeführten ORFs wird unter Verwendung von "Bacterial Artificial Chromosomes", Cosmiden (z. B. Supercos; Stratagene GmbH, Heidelberg) und T7-Expressionssystemen (Stratagene GmbH, Heidelberg; New England Biolabs GmbH, Schwalbach, FRG) vorgenommen. Die Expression rekombinierter Enzyme erfolgt in Escherichia coli-Zellen, die eine konstitutive Expression einer Phosphopantetheinyl-Transferase gewährleisten, welche für die Bildung von Holoenzym-Polyketidsynthasen und -Polypeptidsynthetasen notwendig ist.

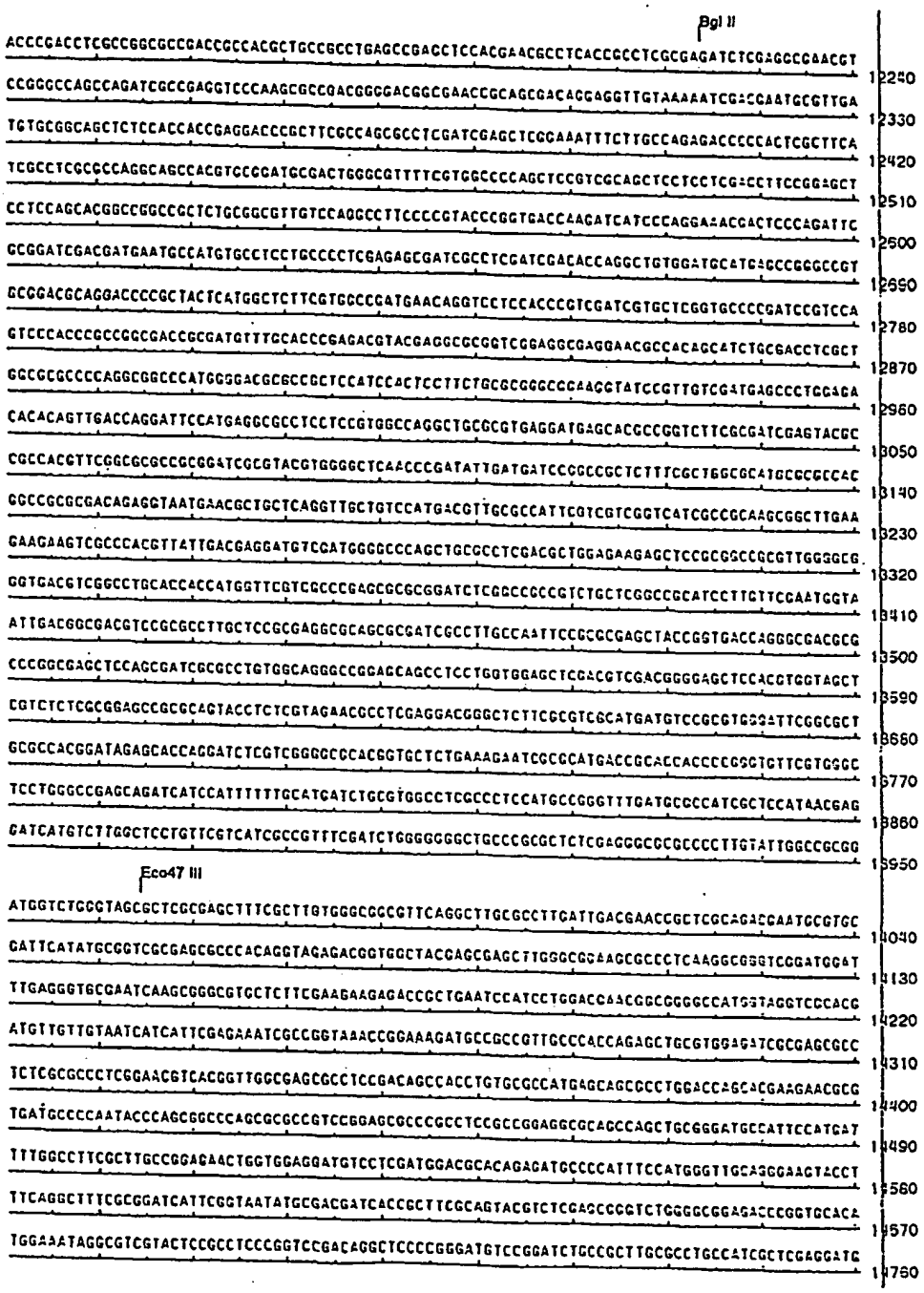

BamH I
Eco47 III

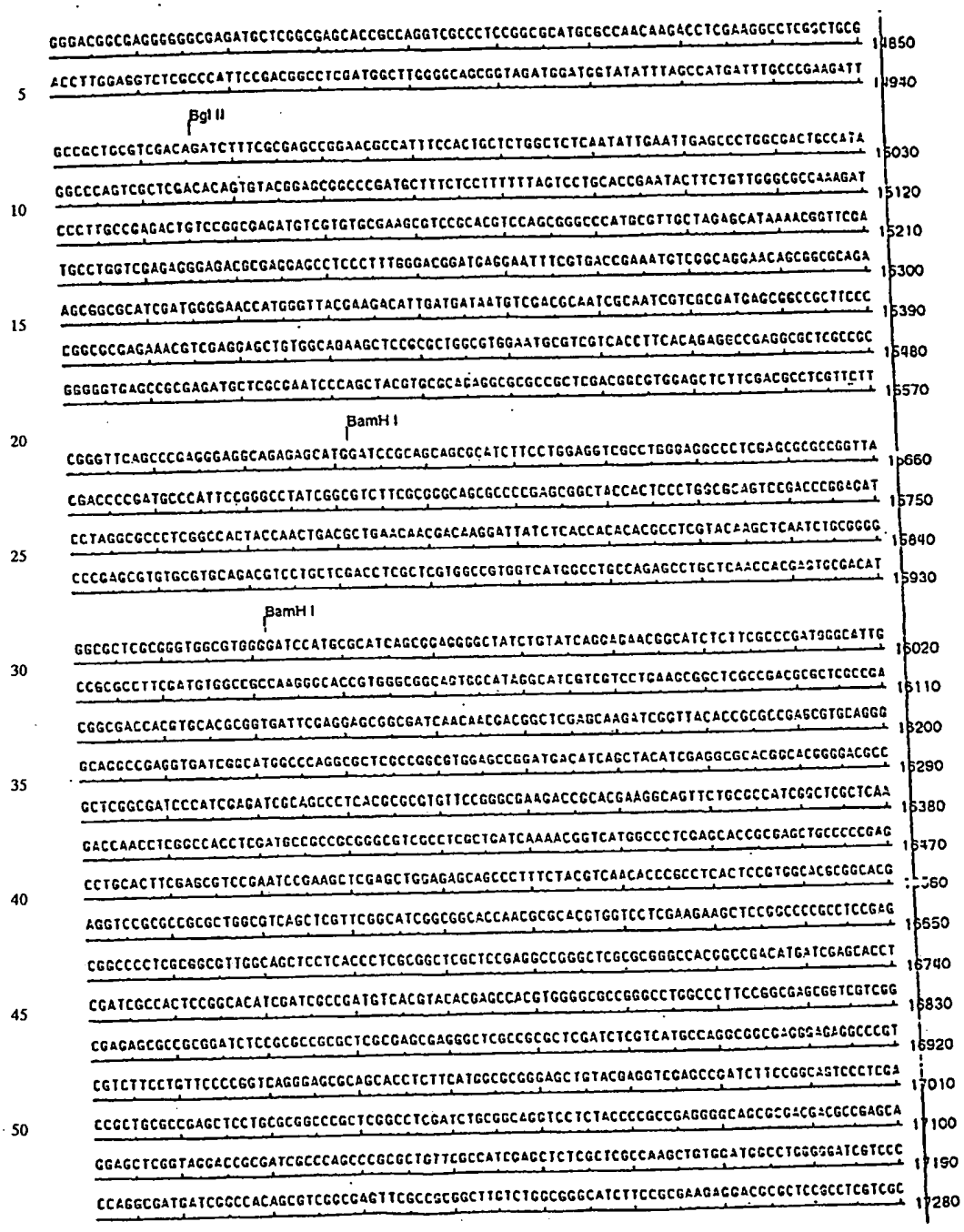

Eco47 III
BamH I
Bgl II
BamH I

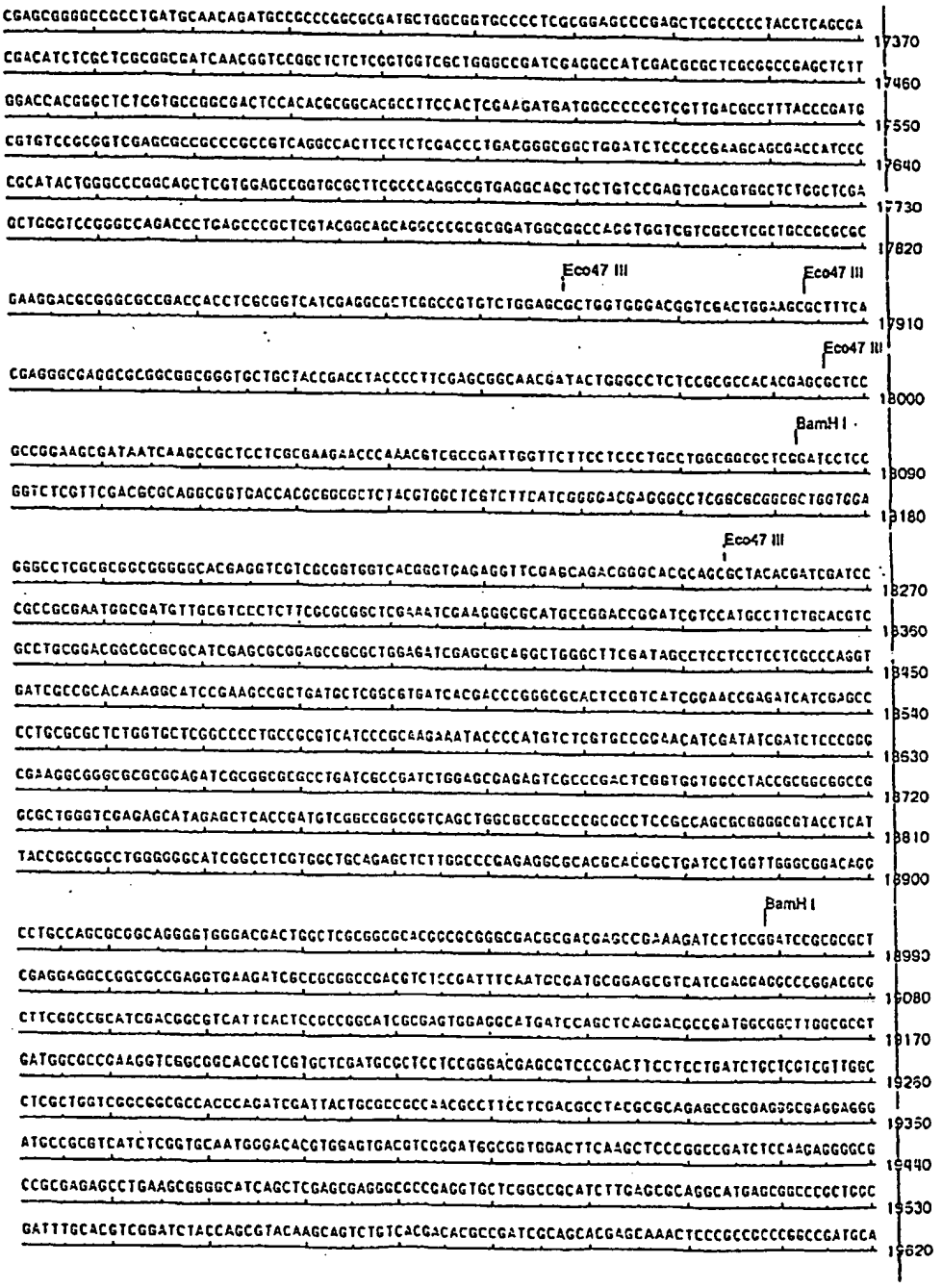

Eco47 III
BamH I
BamH I
BamH I
BamH I

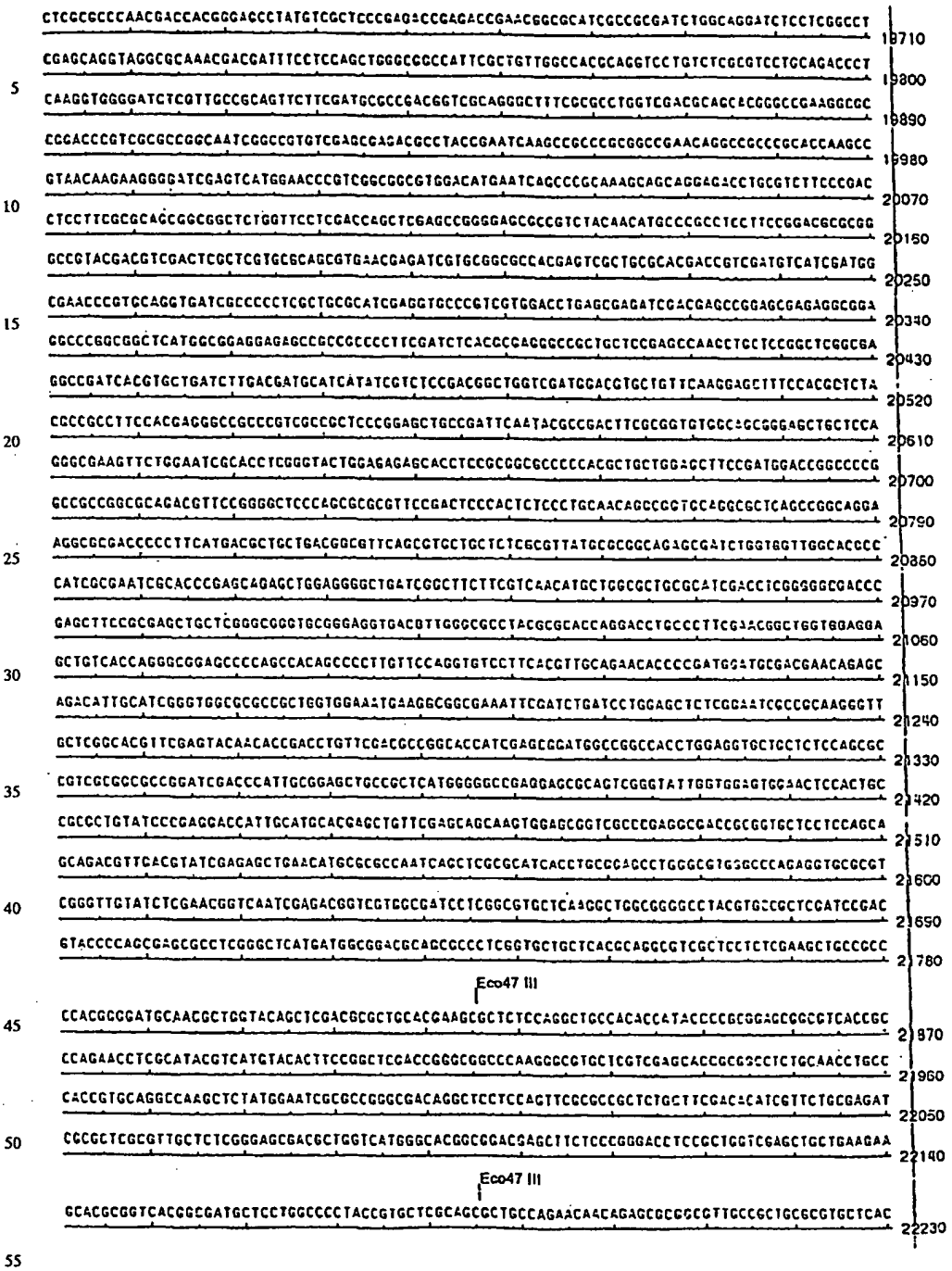

BamH I
BamH I
Eco47 III

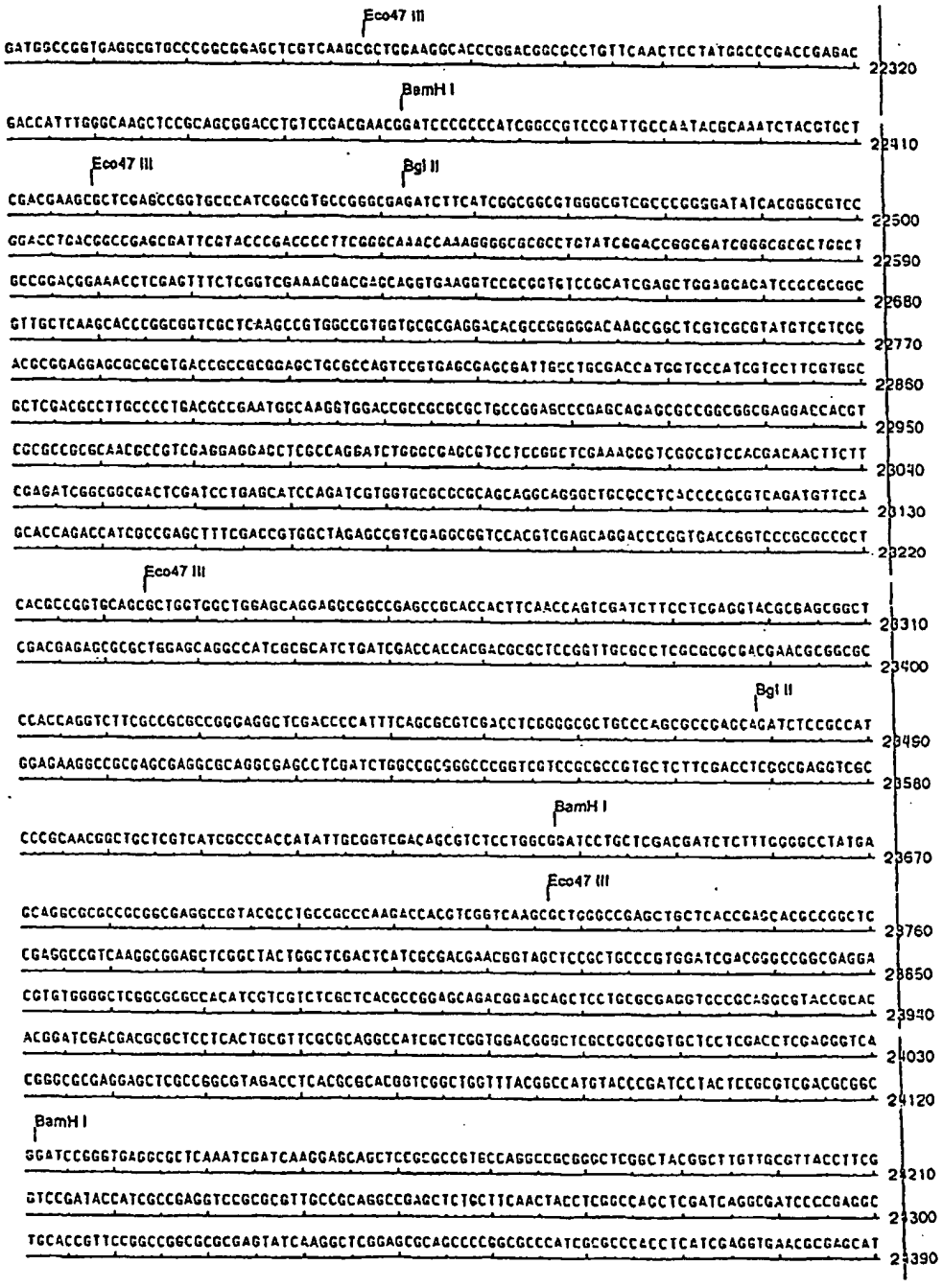

ATGGCGACTCGGATCGTGTACCTGCCGTGGAGACGGGTATGGGACAAAAATACCCTGCCCGACTTGTTGACCTCGTCCAGCAGCGCCTCG
10080
TTGAGGCGATCGAGCTCGCGTTCGATCGACTCTCTCTCTGCCTCGTCCGCCCACCGCATGATGCAAGCGAGCCCCGCAGGCCCTCATGCGA
10170
AAGCAGACCGTACTGAACGGCGTCGGCGCGAGGCGCTCCCAATCGGGATCGGCGTCCACCCACTGGGCGAGCTGCTGCCCCAATCGCAGC
10260
BamH I
Eco47 III
TGCTCCCGGATCCGGGCCGCCAGCCCTTCATGCCCGAAGTAGCGCACGATCATCCAGAGCTTCAGCGCTCGGAAGCGCCGACCGAGCTGG
10350
ATACCCCAGTCCATGTAATTCGTGACGTCGCCCTCGGTGCCGAGGTATTCGGGCACCAGACTGAACGCGCGCTTCAGTCGGTCGGCGTCA
10440
CGCACGTAGAGCACGCTGCAATCCATGGGGGTGAACAGCCACTTGTGAGGGTTCACTACCAGCGAGTCCGCCCCCTCGCAGCCCGCGAGC
10530
ACGTCCCTGTGCTCGGGGACGATCGCGGCCATCCCCGCGTAGGCCGCGTCCACGTGAAGCCATAGCCCGTGCTCCCGGCAAACGCTGACG
10620
ATGGCGGGGATGGGGTCGACGCTCGTCGTGGACGTCGTGCCCACCGTCGCCGCGACGCAGAAGGGTCGGAGGCCGCCCCCGAGGTCCTCC
10710
BamH I
ACGACGGCGGCGCGCAGCGCCTCGGCGACCATGCGGAAGGCCGGATCCGTGGGGATCTTCCGCACCCCCTCCTGCCCGATGCCGAGGGTG
10800
ATGGCTGCCTTCTCGATGGATGAGTCCGCCTGCTCCGACGCGTAGAGTCGCATCCGCCGCTGTCCCGCCATGCCCCGGAGCCGGATCGTC
10890
GGCTCGGCCGAGTCGCGCGCGGCCGCGATCGCGACCATGCTGGCGGTCGACGCGGTGTCCATGATCGCGCCGTGCAAGCCGGCGTCGAGA
10980
TCCAGCATCTGACGCAGCCAGGAGAGGACGAGCTCCTCGAGCTCGGTGGCCGCCGGCGACGTGCGCCATAGCATCACGTTGACGTTGAGG
11070
BamH I
CACGCCGCGAGCAGCTCGCCGAGGATCCCAGGACCAGACGCCGTGTTCGCGAAATACGCGAAGAATCGCGGATGATTCCAGTGCGTGATC
11160
CCCGGCAGAATGATCTGCTCGAAATCGGTGAGCACGGCGTCCATCGGCTCCGGCTCGACGGGCGGGGTCGGGGCCAGCCTGCCCTTCACG
11250
TCGCCGGGGCGGATCGCGGGAAAGACGGGGTATCGATCCGGGTGGCCGAGGTAATCGGCCGCCCAATCGATGATTCTCATACCGATCCGG
11340
BamH I
CGGAACTCCTCCAGATCCATGTCCCCGAGCCGTTCTTTCCGCGGGTCGCTCACGTCAACCTCCTCGCCCTGCCAGGACAGGATCCTCGAG
11430
Eco47 III
GTCCCCTGGCTCCGGCGGTGGAAAGCGCTCCTTGAACGTGAAGGCCCACGGGGTCGGTCCGTAGCGCCGCAGGTGCTCGAGCCGATCCTG
11520
CCCCTCGCGGACGGACGGGATGTGCCCGGCCGGGACCCACCACAGCACGAGGTAATGCGGCTCGAGATGCTCGAACCACCGAGCGCGCTG
11610
TCGCAGGAACGCGGCATGATCCGCGGTGTAGGTGAAGGCGAACAGGTGCTCGATGGAGGTCCATACCGACAGGGTCACGAGGAGCCGCTG
11700
GTCCGGGTACGGACGGATGGACACAGAGTTCCCCTCGGCCGTCTGCAGGCGCCACACGAACCCCTCGCTCCGATCGGCCAGATGGTTGAT
11790
BamH I
ATGGTCGAGCCCCTGGACGAAGCCCTCCATGATCGGATCCTCCAGCGGAGCGCGAATACATGCGAAGTTGTATTGCGCGATGTGGTGCCG
11880
Eco47 III
ATGCTCCGACATGTCGCTTTCCATCTCCAGCTCCCGCTCACCAATCCCAGCGCTGCTCCGGGGAGCTCATCAGGGCAGACGCGACATCGA
11970
TCCCGAAGCTCCGCCGCATCCCCTCGACGAAGGCGGCCTGGACCGCTTCGGCGACGGATCGGCCTGCCTCCGGCAAGACCTCCGAGACAA
12050
AGAAGAACCGCCTCGTGGAAGGGACAATCTTGCCCCGCTCCGCCTGGCGCCATACGAAGTGCCTCGTCACCAGTCCCTCCGCGTCGGCAT
12150

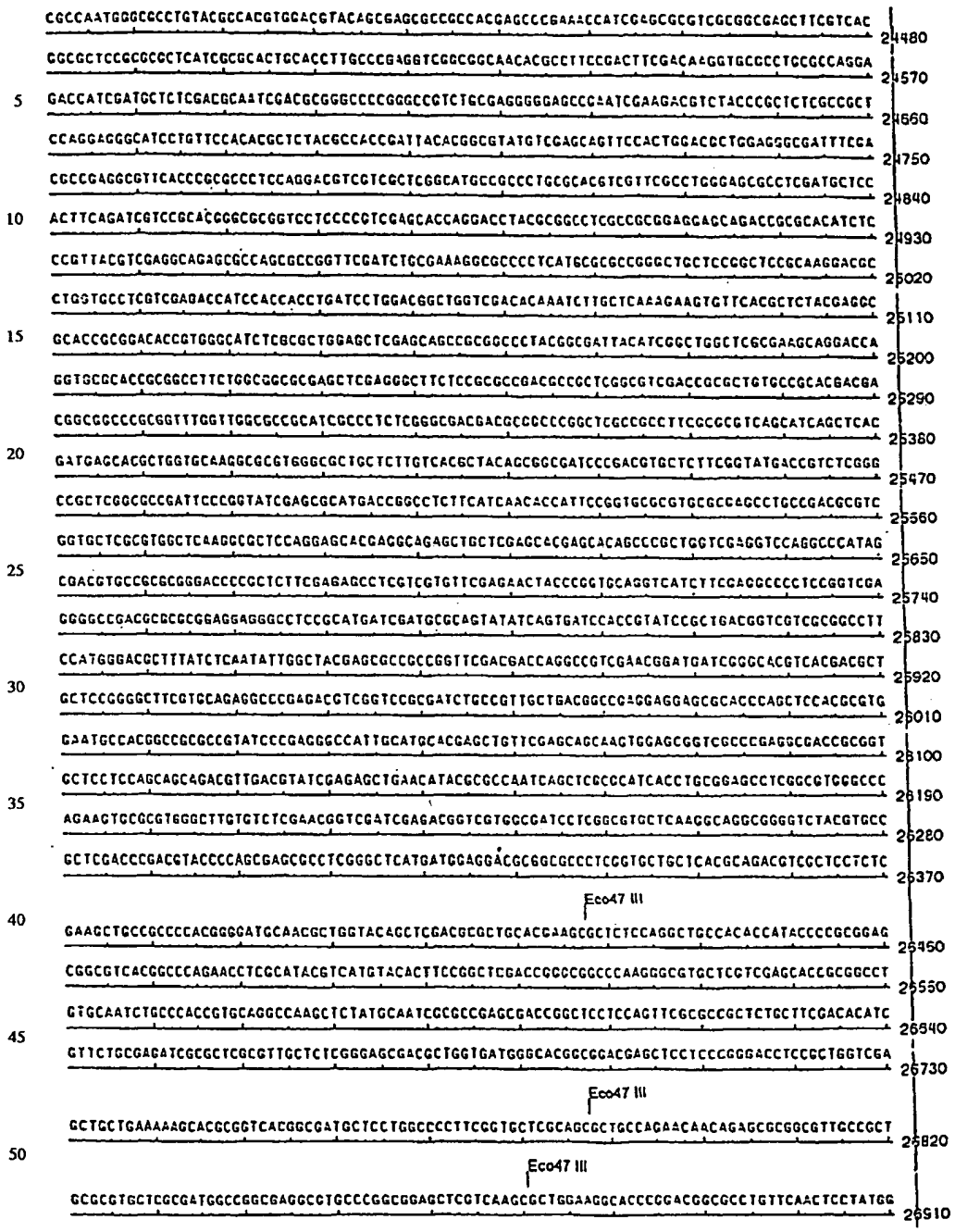

Bgl II
Eco47 III

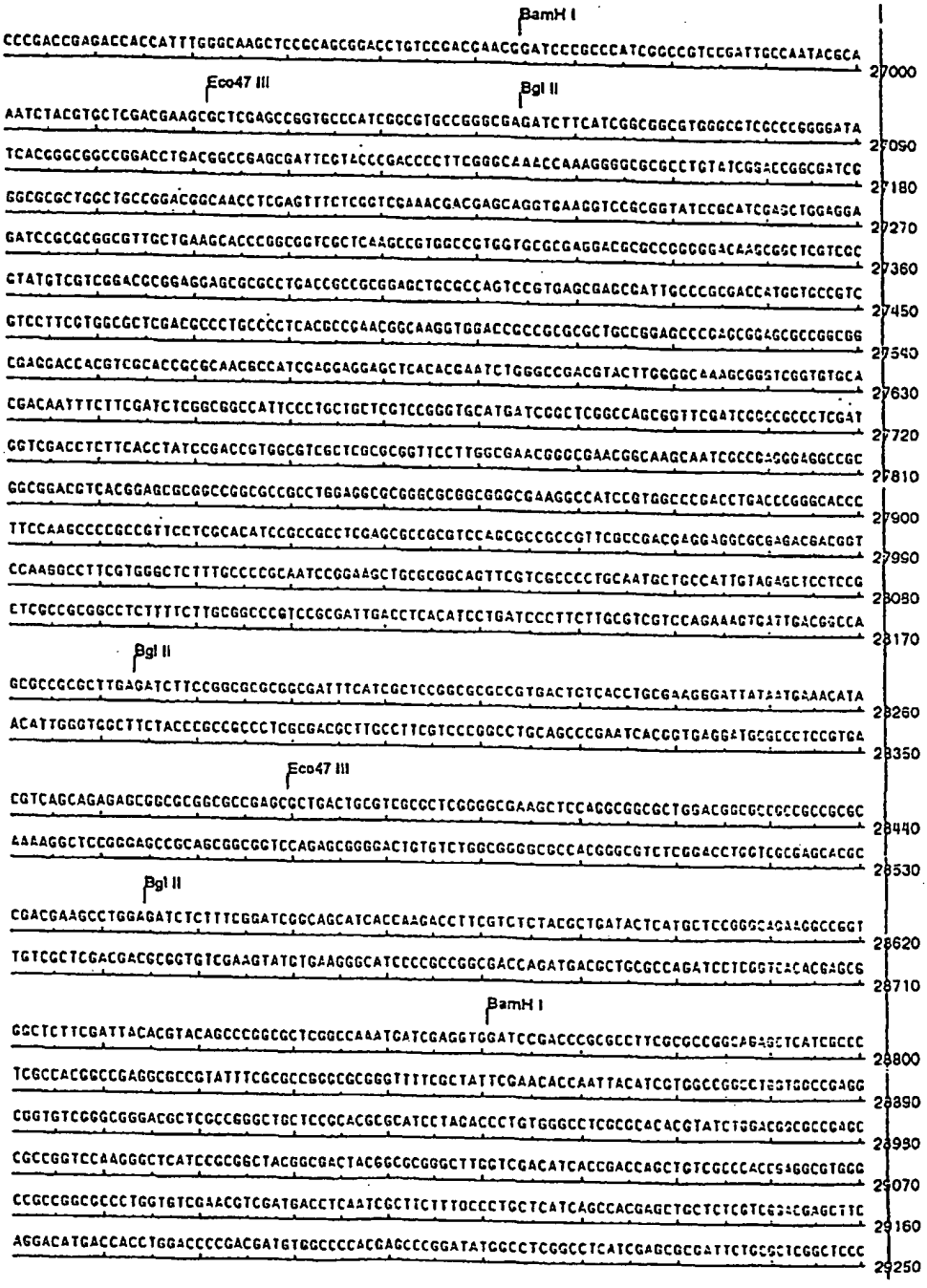

Bgl II
BamH I
BamH I

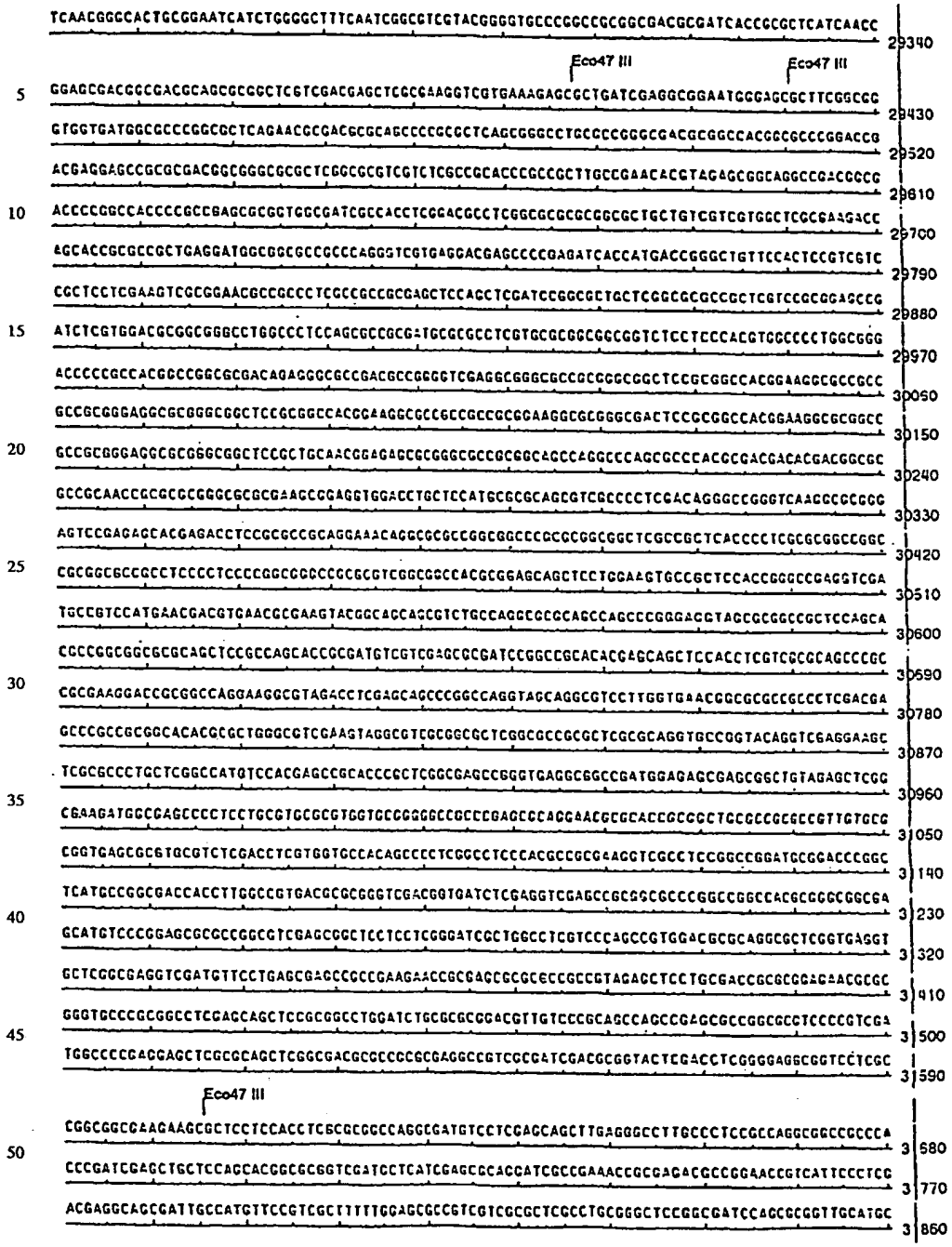

Eco47 III
Eco47 III
Eco47 III
BamHI
Eco47 III
BamHI

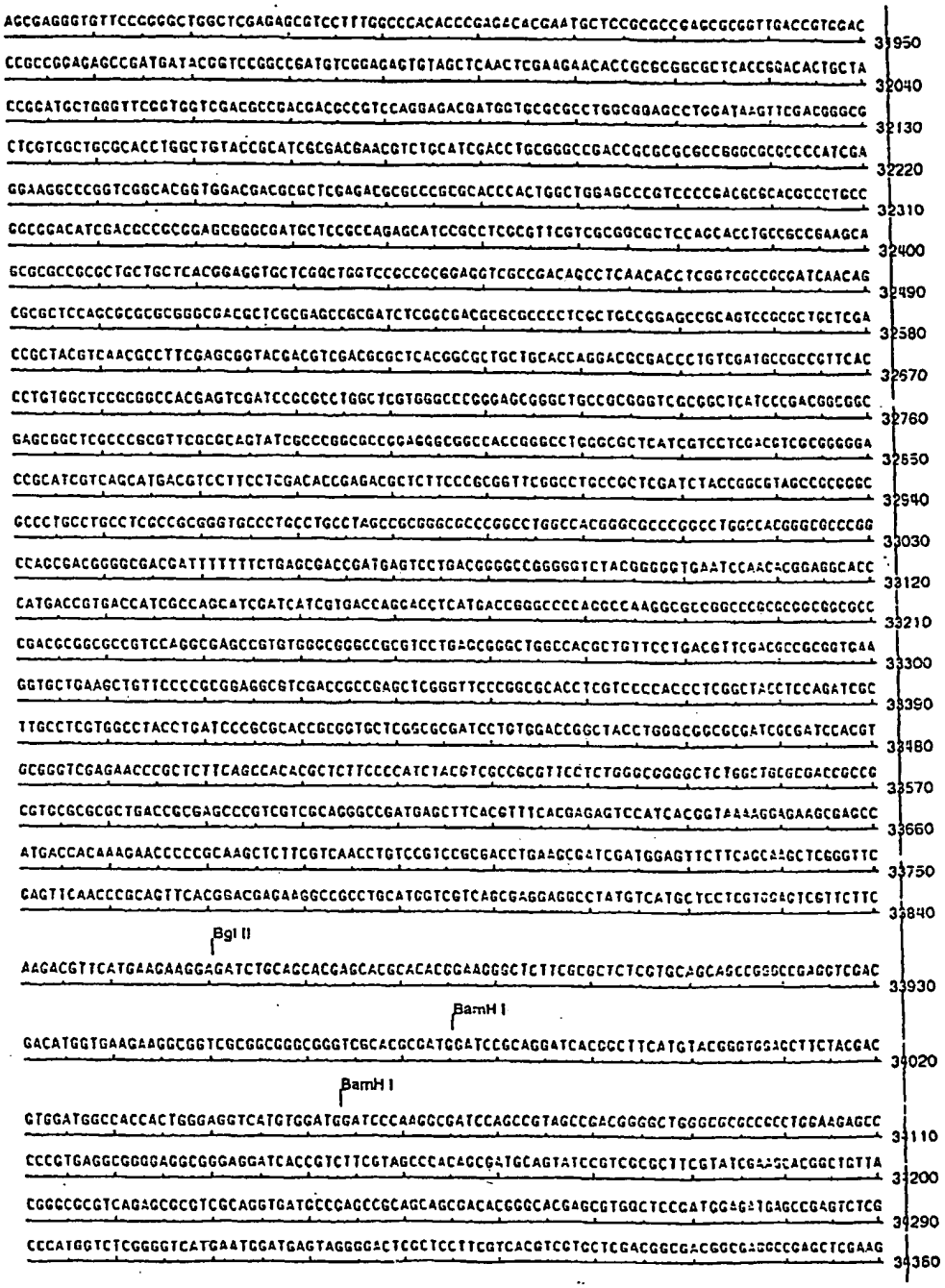

CTCGCGCCCAACGACCACGGGAGCCTATGTCGCTCCCGAGACCGAGACCGAACGGCGCATCGCCGCGATCTGGCAGGATCTCCTCGGCCT
19710
CGAGCAGGTAGGCGCAAACGACGATTTCCTCCAGCTGGGCGCCCATTCGCTGTTGGCCACGCAGGTCCTGTCTCGCGTCCTGCAGACCCT
19800
CAAGGTGGGGATCTCGTTGCCGCAGTTCTTCGATGCGCCGACGGTCGCAGGGCTTTCGCGCCTGGTCGACGCAGCACGGGCCGAAGGCGC
19890
CGGACCCGTCGCGCCGGCAATCGGCCGTGTCGAGCGAGACGCCTACCGAATCAAGCCGCCCGCGGCCGAACAGGCCGCCCGCACCAAGCC
19980
GTAACAAGAAGGGGATCGAGTCATGGAACCCGTCGGCGGCGTGGACATGAATCAGCCCGCAAAGCAGCAGGAGACCTGCGTCTTCCCGAC
20070
CTCCTTCGCGCAGCGGCGGCTCTGGTTCCTCGACCAGCTCGAGCCGGGGAGCGCCGTCTACAACATGCCCGCCTCCTTCCGGACGCGCGG
20160
GCCGTACGACGTCGACTCGCTCGTGCGCAGCGTGAACGAGATCGTGCGGCGCCACGAGTCGCTGCGCACGACCGTCGATGTCATCGATGG
20250
CGAACCCGTGCAGGTGATCGCCCCCTCGCTGCGCATCGAGGTGCCCGTCGTGGACCTGAGCGAGATCGACGAGCCGGAGCGAGAGGCGGA
20340
GGCCCGGCGGCTCATGGCGGAGGAGAGCCGCCGCCCCTTCGATCTCACGCGAGGGCCGCTGCTCCGAGCCAACCTGCTCCGGCTCGGCGA
20430
GGCCGATCACGTGCTGATCTTGACGATGCATCATATCGTCTCCGACGGCTGGTCGATGGACGTGCTGTTCAAGGAGCTTTCCACGCTCTA
20520
CGCCGCCTTCCACGAGGGCCGCCCGTCGCCGCTCCCGGAGCTGCCGATTCAATACGCCGACTTCGCGGTGTGGCAGCGGGAGCTGCTCCA
20610
GGGCGAAGTTCTGGAATCGCACCTCGGGTACTGGAGAGAGCACCTCCGCGGCGCCCCCACGCTGCTGGAGCTTCCGATGGACCGGCCCCG
20700
GCCGCCGGCGCAGACGTTCCGGGGCTCCCAGCGCGCGTTCCGACTCCCACTCTCCCTGCAACAGCCGGTGCAGGCGCTCAGCCGGCAGGA
20790
AGGCGCGACCCCCTTCATGACGCTGCTGACGGCGTTCAGCGTCGTGCTCTCGCGTTATGCGCGGCAGAGCGATCTGGTGGTTGGCACCCC
20880
CATCGCGAATCGCACCCGAGCAGAGCTGGAGGGGCTGATCGGCTTCTTCGTCAACATGCTGGCGCTGCGCATCGACCTCGGGGGCGACCC
20970
GAGCTTCCGCGAGCTGCTCGGGCGGGTGCGGGAGGTGACGTTGGGCGCCTACGCGCACCAGGACCTGCCCTTCGAACGGCTGGTGGAGGA
21060
GCTGTCACCAGGGCGGAGCCCCAGCCACAGCCCCTTGTTCCAGGTGTCCTTCACGTTGCAGAACACCCCGATGGATGCGACGAACAGAGC
21150
AGACATTGCATCGGGTGGCGCGCCGCTGGTGGAAATGAAGGCGGCGAAATTCGATCTGATCCTGGAGCTCTCGGAATCGCCGCAAGGGTT
21240
GCTCGGCACGTTCGAGTACAACACCGACCTGTTCGACGCCGGCACCATCGAGCGGATGGCCGGCCACCTGGAGGTGCTCCTCTCCAGCGC
21330
CGTCGCGCCGCCGGATCGACCCATTGCGGAGCTGCCGCTCATGGGGGCCGAGGAGCGCAGTCGGGTATTGGTGGAGTGGAACTCCACTGC
21420
CGCCCTGTATCCCGAGGACCATTGCATGCACGAGCTGTTCGAGCAGCAAGTGGAGCGGTCGCCCGAGGCGACCGCGGTGCTCCTCCAGCA
21510
GCAGACGTTCACGTATCGAGAGCTGAACATGCGCGCCAATCAGCTCGCGCATCACCTGCCGAGCCTGGGCGTGGGCCCAGAGGTGCGCGT
21600
CGGGTTGTATCTCGAACGGTCAATCGAGACGGTCGTGCCGATCCTCGGCGTGCTCAAGGCTGGCGGGGCCTACGTGCCGCTCGATCCGAC
21690
GTACCCCAGCGAGCGCCTCGGGCTCATGATGGCGGACGCAGCGCCCTCGGTGCTGCTCACGCAGGCGTCGCTCCTCTCGAAGCTGCCGCC
21780
Eco47 III
CCACGCGGATGCAACGCTGGTACAGCTCGACGCGCTGCACGAAGCGCTCTCCAGGCTGCCACACCATACCCCGCGGAGCGGCGTCACCGC
21870
CCAGAACCTCGCATACGTCATGTACACTTCCGGCTCGACCGGGCGGCCCAAGGGCGTGCTCGTCGAGCACCGCGGCCTCTGCAACCTGCC
21960
CACCGTGCAGGCCAAGCTCTATGGAATCGCGCCGGGCGACAGGCTCCTCCAGTTCGCGCCGCTCTGCTTCGACACATCGTTCTGCGAGAT
22050
CGCGCTCGCGTTGCTCTCGGGAGCGACGCTGGTCATGGGCACGGCGGACGAGCTTCTCCCGGGACCTCCGCTGGTCGAGCTGCTGAAGAA
22140
Eco47 III
GCACGCGGTCACGGCGATGCTCCTGGCCCCTACCGTGCTCGCAGCGCTGCCAGAACAACAGAGCGCGGCGTTGCCGCTGCGCGTGCTCAC
22230

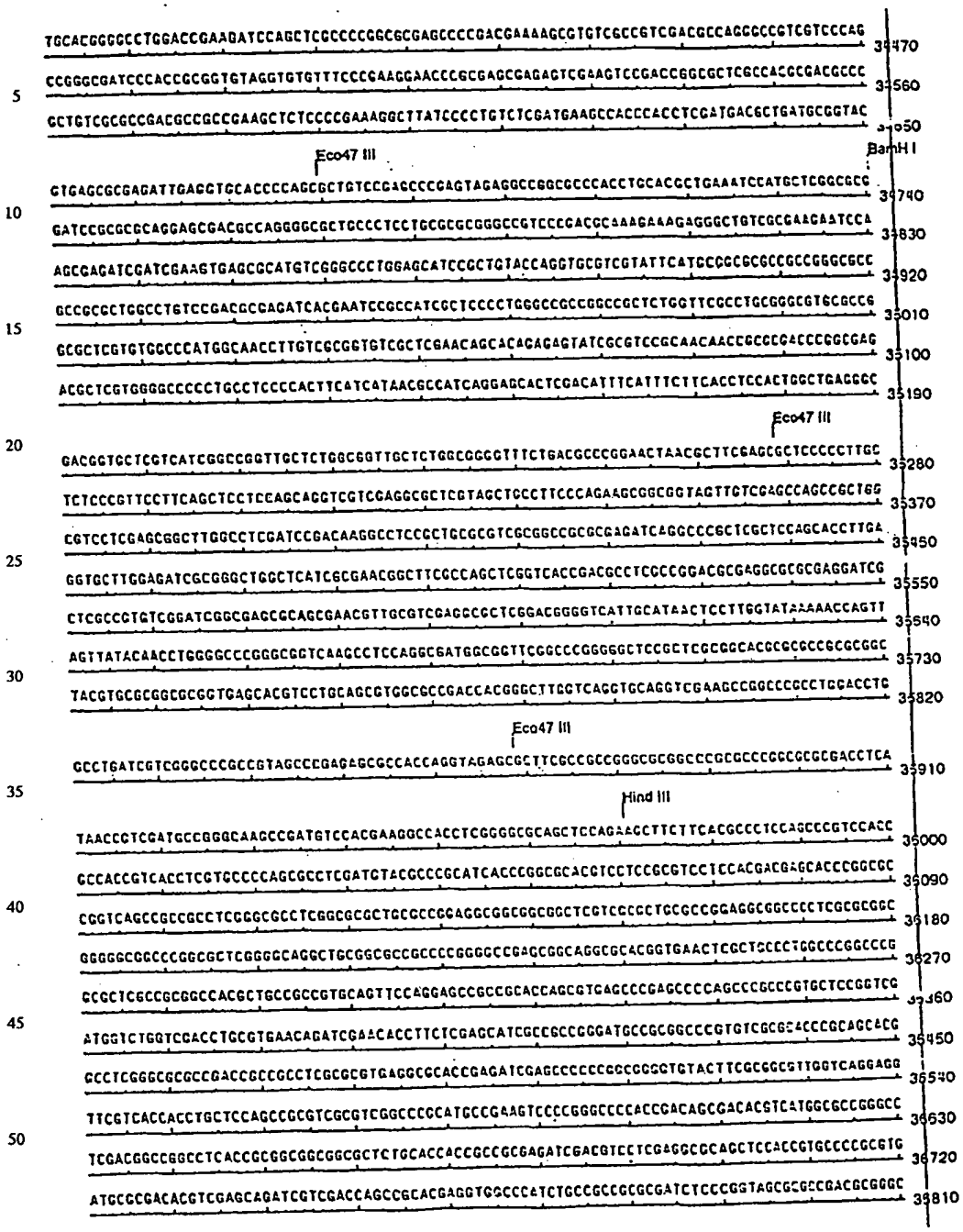

Eco47 III
BamH I
Eco47 III
Bgl II
Eco47 III
Bgl II
BamH I
Eco47 III
BamH I

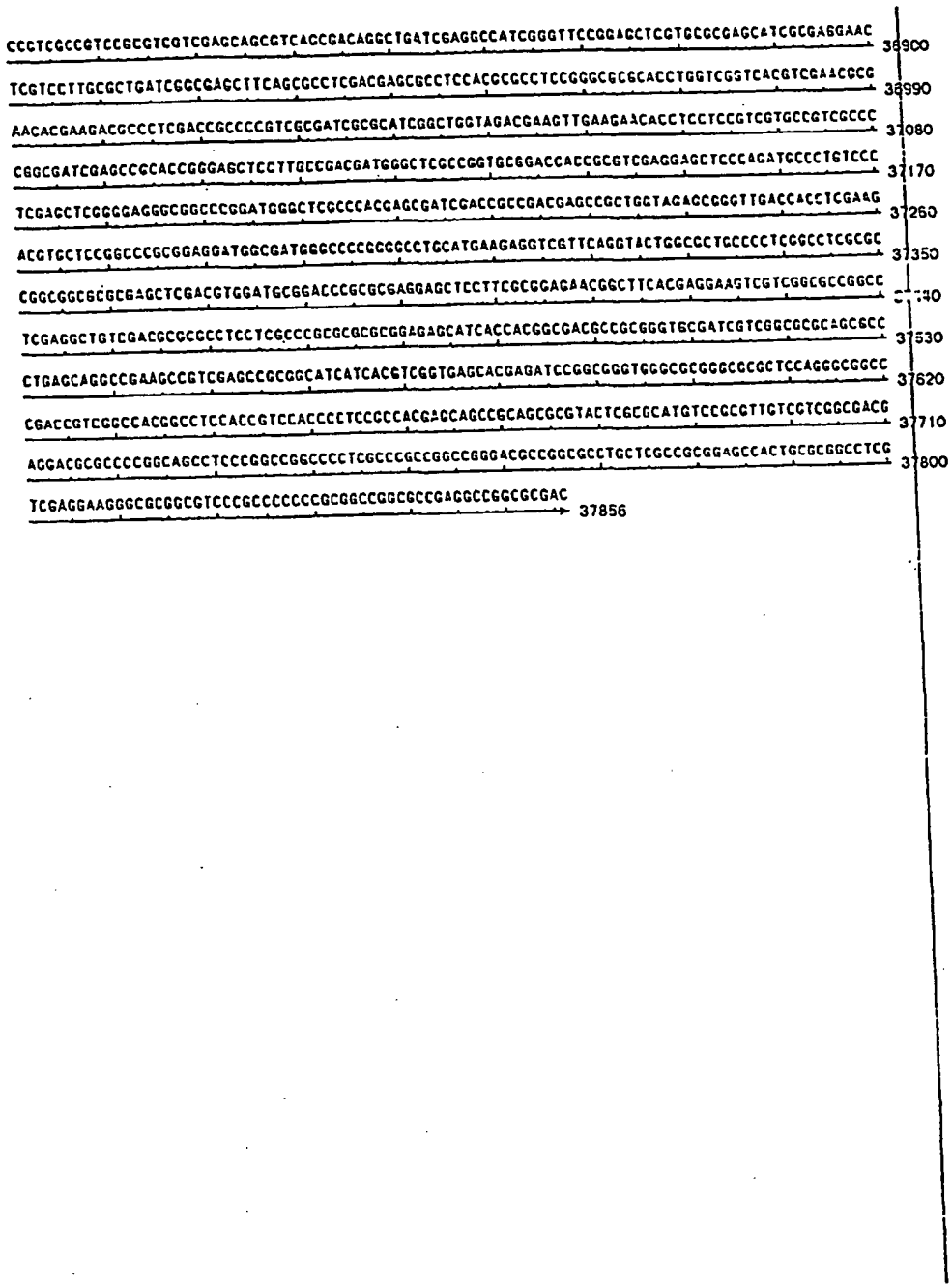

24480
24570
24660
24750
24840
24930
25020
25110
25200
25290
25380
25470
25560
25650
25740
25830
25920
26010
26100
26190
26280
26370
Eco47 III
26460
26550
26640
26730
Eco47 III
26820
Eco47 III
26910


BamH I
Eco47 III
Bgl II
Bgl II
Eco47 III
Bgl II
BamH I


Eco47 III
Eco47 III
Eco47 III


AGCGAGGGTGTTCCCGGGCTGGCTCGAGAGCGTCCTTTGGCCCACACCCGAGACACGAATGCTCCGCGCCCGAGCGCGGTTGACCGTCGAC
31950
CCGCCGGAGAGCCGATGATACGGTCCGGCCGATGTCGGAGAGTGTAGCTCAACTCGAAGAACACCGCGCGGCGCTCACCGGACACTGCTA
32040
CCGGATGCTGGGTTCGGTGGTCGACGCCGACGACGCCGTCCAGGAGACGATGGTGCGCGCCTGGCGGAGCCTGGATAAGTTCGACGGGCG
32130
CTCGTCGCTGCGCACCTGGCTGTACCGCATCGCGACGAACGTCTGCATCGACCTGCGGGCCGACCGCGCGCGCCGGGCGCGCCCCATCGA
32220
GGAAGGCCCGGTCGGCACGGTGGACGACGCGCTCGAGACGCGCCCGCGCACCCACTGGCTGGAGCCCGTCCCCGACGCGCACGCCCTGCC
32310
GGCCGACATCGACGCCGCGGAGCGGGCGATGCTCCGCCAGAGCATCCGCCTCGCGTTCGTCGCGGCGCTCCAGCACCTGCCGCCGAAGCA
32400
GCGCGCCGCGCTGCTGCTCACGGAGGTGCTCGGCTGGTCCGCCGCGGAGGTCGCCGACAGCCTCAACACCTCGGTCGCCGCGATCAACAG
32490
CGCGCTCCAGCGCGCGCGGGCGACGCTCGCGAGCCGCGATCTCGCCGACGCGCGCCCCTCGCTGCCGGAGCCGCAGTCCGCGCTGCTCGA
32580
CCGCTACGTCAACGCCTTCGAGCGGTACGACGTCGACGCGCTCACGGCGCTGCTGCACCAGGACGCGACCCTGTCGATGCCGCCGTTCAC
32670
CCTGTGGCTCCGCGGCCACGAGTCGATCCGCGCCTGGCTCGTGGGCCCGGGAGCGGGCTGCCGCGGGTCGCGGCTCATCCCGACGGCGGC
32760
GAGCGGCTCGCCCGCGTTCGCGCAGTATCGCCCGGCGCCGGAGGGCGGCCACCGGGCCTGGGCGCTCATCGTCCTCGACGTCGCGGGGGA
32850
CCGCATCGTCAGCATGACGTCCTTCCTCGACACCGAGACGCTCTTCCCGCGGTTCGGCCTGCCGCTCGATCTACCGGCGTAGCCGCGGGC
32940
GCCCTGCCTGCCTCCCCGCGGGTGCCCTGCCTGCCTAGCCGCGGGCGCCCGGCCTGGCCACGGGCGCCCGGCCTGGCCACGGGCGCCCGG
33030
CCAGCGACGGGGCGACGATTTTTTTCTGAGCGACCGATGAGTCCTGACGGGGCCGGGGGTCTACGGGGGTGAATCCAACACGGAGGCACC
33120
CATGACCGTGACCATCGCCAGCATCGATCATCGTGACCAGGACCTCATGACCGGGCCCCACGCCAAGGCGCCGGCCCGCCGCGGCGGCGCC
33210
CGACGCGGCGCCGTCCAGGCGAGCCGTGTGGGCGGGCCGCGTCCTGAGCGGGCTGGCCACGCTGTTCCTGACGTTCGACGCCGCGGTGAA
33300
GGTGCTGAAGCTGTTCCCCGCGGAGGCGTCGACCGCCGAGCTCGGGTTCCCGGCGCACCTCGTCCCCACCCTCGGCTACCTCCAGATCGC
33390
TTGCCTCGTGGCCTACCTGATCCCGCGCACCGCGGTGCTCGGCGCGATCCTGTGGACCGGCTACCTGGGCGGCGCGATCGCGATCCACGT
33480
GCGGGTCGAGAACCCGCTCTTCAGCCACACGCTCTTCCCCATCTACGTCGCCGCGTTCCTCTGGGCGGGGCTCTGGCTGCGCGACCGCCG
33570
CGTCGCGCGCTGACCGCGAGCCCGTCGTCGCAGGGCCGATGAGCTTCACGTTTCACGAGAGTCCATCACGGTAAAAGGAGAAGCGAGCC
33660
ATGACCACAAAGAACCCCCGCAAGCTCTTCGTCAACCTGTCCGTCCGCGACCTGAAGCGATCGATGGAGTTCTTCAGCAAGCTCGGGTTC
33750
GAGTTCAACCCGCAGTTCACGGACGAGAAGGCCGCCTGCATCGTCGTCAGCGAGGAGGCCTATGTCATGCTCCTCGTGGACTCGTTCTTC
33840
BglII
AAGACGTTCATGAAGAAGGAGATCTGCAGCACGAGCACGCACACGGAAGGGCTCTTCGCGCTCTCGTGCAGCAGCCGGGCCGAGGTCGAC
33930
BamHI
GACATGGTGAAGAAGGCGGTCGCGGCGGCGGGTCGCACGCGATGGATCCGCAGGATCACCGCTTCATGTACGGGTGGAGCTTCTACGAC
34020
BamHI
GTGGATGGCCACCACTGGGAGGTCATGTGGATGGATCCCAAGGCGATCCAGCCGTAGCCGACGGGGCTGGGCGCGCCCCCTGGAAGAGCC
34110
CCCGTGAGGCGGGGAGGCGGGAGGATCACCGTCTTCGTAGCCCACAGCGATGCAGTATCCGTCGCGCTTCGTATCGAAGCACGGCTGTTA
34200
CGGGCGCGTCAGAGCGCGTCGCAGGTGATCCCAGCCGCAGCAGCGACACGGGCACGAGCGTGGCTCCGATGGAGTGAGCCGAGTCTCG
34290
CCCATGGTCTCGGGGTCATGAATGGATGAGTAGGGGACTCGCTCCTTCGTCACGTCGTCCTCGACGGCGACGGCGAGGCCGAGCTCGAAG
34380


Eco47 III
BamH I
Eco47 III
Eco47 III
Hind III


CCGTCGCCGTCCGCGTCGTCGAGCAGCGTCACCGACAGGCTGATCGAGGCCATCGGGTTCCGGAGCTCGTGCGCGAGCATCGCGAGGAAC
38900
TCGTCCTTGCGCTGATCGGCGAGCTTCAGCGCCTCGACGAGCGCCTCCACGCCCCTCCGGGCGCGCACCTGGTCGGTCACGTCGAACCCG
38990
AACACGAAGACGCCCTCGACCGCCCCGTCGCGATCGCGCATCGGCTGGTAGACGAAGTTGAAGAACACCTCCTCCGTCGTGCCGTCGCCC
37080
CGGCGATCGAGCCCGCACCGGGAGCTCCTTGCCGACGATGGGCTCGCCGGTGCGGACCACCGCGTCGAGGAGCTCCCAGATGCCCTGTCCC
37170
TCGAGCTCGGGGAGGGCGGCCCGGATGGGCTCGCCCACGAGCGATCGACCGCCGACGAGCCGCTGGTAGACCGGGTTGACCACCTCGAAG
37260
ACGTGCTCCGGCCCGCGGAGGATGGCGATGGGCCCCGGGGCCTGCATGAAGAGGTCGTTCAGGTACTGGCCCTGCCCCTCGGCCTCGCGC
37350
CGGCGGCGCGCGAGCTCGACGTGGATGCGGACCCGCGCGAGGAGCTCCTTCGCGGAGAACGGCTTCACGAGGAAGTCGTCGGCGCCGGCC
37440
TCGAGGCTGTCGACGCGCGCCTCCTCGCCCGCGCGCGCGCGGAGAGCATCACCACGGCGACGCCGCGGGTGCGATCGTCGGCGCGCAGCGCC
37530
CTGAGCAGGCCGAAGCCGTCGAGCCGCGGCATCATCACGTCGGTGAGCACGAGATCCGGCGGGTGGGCGCGGGCGCCGCTCCAGGGCGGCC
37620
CGACCGTCGGCCACGGCCTCCACCGTCCACCCCTCCCCCACGAGCAGCCGCAGCGCGTACTCGCGCATGTCCGCGTTGTCGTCCGGCGACG
37710
AGGACGCGCCCCGGCAGCCTCCCGGCCGGCCCCTCGCCCGCCGGCCGGGACGCCCGGCGCCTGCTCGCCGCGGAGCCACTGCGCGGCCTCG
37800
TCGAGGAAGGGCGCGGCGTCCCGCCCCCCCGCGGCCGGCGCCGAGGCCGGCGCGAC
37856

SEQUENCE LISTING

<110> Gesellschaft für Biotechnologische Forschung GmbH

<120> DNA-Sequenzen für die enzymatische Synthese von Polyketid- oder Heteropolyketidverbindungen

<130> DNA-Seq zur enzym Syn von Polyketid .

<140> DE 198 46 493.2

<141> 1998-10-09

<160> 1

<170> PatentIn Ver. 2.1

<210> 1

<211> 37856

<212> DNA

<213> Sorangium cellulosum

<400> 1

ggatcgcggc gccctcgcgc tgctcctcga gcgtgcggag gaactcccac gccaggcgcg 60
acttgccgag gccaggcgcg cccaccacca ccaccgcgtt cgcggagggc tcgtcgacgc 120
aatggcgcca ctcggtcgcg agctgcgaga gctcgcgctc ccgccccacg cagggcgtcg 180
gcttgccgag gagccgtggg acggcatccg gctcctcctt cgggccgcga agccagcacc 240
ctccgggccc ctgtaccgtc tcgaagcggc tcgcgagcag gctggccgtc gcgtcgtcga 300
gccggatctc cggcggcgac aggccatctc gcccggcgat gagctgcgcg acccgatcga 360
ccagctcgcc gaccggcagc ctcgcctcga cctcggccag ccctgtcgcg acggacacgg 420
gcacgcctcc gagcgccgcc cgcagcgcga gggcgcagtg ggccgcccgt gtggcgagat 480
ccgtgggcga ctcggcgccg gacagcgcga cgagccacca gcgcgcttgc agccgatcga 540
ggcgcccgcc gtggcgcgcc gcgatgtccc gcagcgcctc ggcccgcgcg gcgccgtcgt 600
cctccgagag cgtggcgccg gcctcggcgc cgccgtcttc ggccaggatg acgcacatca 660
ccttgcgctc ggccgtcgtg atcgcctcgc ccggcgcggc cggcgccgcg accgcgctcg 720
ccccgatcga gagcccctcg ccggccacgg cggcgagctc cgccgcggcg gcggcgccgt 780
cgcgcggccg ctctcccgcg ttcttcgcca gcatccgcgc caccaggcgc tcgagcggct 840
ccgggatacc gtcgcggagc tccccgagcc gcggcggctc ttccaggacg acccgcatca 900
ggagcgcgag cgcgctgttg ccgaggaacg gcgggcgccc cgcgaggcac tggaacagca 960
cgcacccgag cgcgaacacg tcggcccggg cgtcgaccgg cgcgtcgccg cgcacctgct 1020
cgggcgctat gtacccgggc gtgccgagca cggccccggg cgacgtgagg gtcggcgcga 1080
gccggaggtg gcgcgcgatg ccgaagtcga gcagcgtgac gcgctcgacc gcgccgccca 1140
cgagcatcag gttgctcggc ttgaggtcgc ggtgaacgac gccgagccag tggatcgcgc 1200
cgagcgtcgt ggccacgcgc gcggccagcg ccacgctctc ggccagcgtg agcggcgccc 1260
cggcgagccg ctcctccagg gtcacgccgt cgagccactc catggccagg tacggccgcc 1320
ctgcgccggt caccccgtgc gccacgtact gcaccacgcc gggcagccgg agcgtcacga 1380
gcgcctccgc ctcccgcgcg aaccggcgca ggtcgttggc gctcgcgccc tgcaagacct 1440
tgagcgcgac cgcctgcccg gacacccggt cgcgcgcccg gtacacgtcc cccatcccgc 1500
cggagacggc gagccgctcg atctcgaaac gatcctcgat cacatccgct gcgcgcatgg 1560

cggtgccaat gtactccgcg cgagcctcgg gccccgcgc gtaagtgcgg ccctgcgccc 1620
ggttgaacgc cagcccgagc gtgaccgcct cgcgctcggg atccacggcc gccggatcgg 1680
tccacgcctc gacgagcgcc tgcgttgaac aacccgccac cgggcgcacg cagccggcat 1740
cgccgcgctg gccacccggc gctgccgccc ttaggctcac ctccgcgatg ccccgctggt 1800
tcaacacggc aggtccctgc aacccggccg atcactacat gctcccggcc gaggagcgct 1860
tgcccgcagt gcgcgatctg gtcgatcgca aggcctactt cgtcctgcac gccccgcggc 1920
agatcggcaa gacgacctcg ctgcgcacgc tcgcccagga tctcacggcc gaagggcgct 1980
acgtggccgt cctcgtctcg gcggaggtcg gcgccccctt ctctgacgat cccggcgcgg 2040
ccgagctcgc gatgctcgca gaatggcgcg gcaccgccgg cgcgcagctc cccgccgatc 2100
tgcggccgcc accgttcccc gatgcgcccg ccggtcagcg catcgggggc gccctgcgcg 2160
cctgggctca ggccgcgccg cgcccgctcg tcgtcttcct cgacgaggcc gacgccctgc 2220
gcgacgcgac gctcgtctcc ctattgcgcc agatccgcag cggctatccc gaccgcccgc 2280
gtgacttccc gcacgcgctc gccctcgtcg gcctgcgcga cgtgcgcgac tacaaggtcg 2340
cgtcggtcga cagcggcagg ctcgggacgt cgagcccctt caacatcaag gtcgagtcgc 2400
tcacgctgcg caacttcacc cgcgacgagg tcgcaacact ctacgctcag cacacggccg 2460
agaccggtca ggtcttccgg ccggacgccg tggaccgcgc cttcgagctc acccagggcc 2520
agccgtggct cgccaacgcg ctcgcccgcc agctcgtcga ggtcctcgtc aaggaccgcg 2580
cgcaacccat cacgtctgcg aacgtcgatc gcgccaagga aatcctcatc gagcggcagg 2640
acacacacct cgacagcctg gtggatcggc tgcgcgagcc gcgcatccgc gcggtgatcg 2700
agccgatgct cgccggcacc gcgttgccga gcgtgccccc cgacgacctt cgtttcgcga 2760
tcgacctcgg cctcgtgcgc atgaccgcgg agggcggcct cgacgtcgcc aaccccatct 2820
accgcgagat catcgtccgc gagctcgcgt tcccgatccg cgcctcactc ccccagatca 2880
aggccacgtg gctcacgcag gacggccgcc tcgacgcgga ccgcctgctc gacgccttcc 2940
tctccttctg gcgccagcac ggcgagccgc tcctcggcgc cgcgccctac catgagatcg 3000
ccccgcacct cgtggtgatg gccttcctcc accgcgtggt gaacggcggt ggcaccgtcg 3060
agcgcgagta cgccatcggc cggggcagga tggatctctg cgttcgttac gcgggcgaga 3120
cgctcgcgat cgagctcaag gtctggcgag acggccgccc cgatcccgtc gccgaggggc 3180
tcgcccagct cgacgagtac ctggccggcc tgggcctcga tcgcggatgg ctcatcctct 3240
tcgaccagcg ctccggacag ccccccatcg ccgagcgcac gcgccgcgag cgcgcgctct 3300
ccccgccggg ccgcgaggtc gccgtcattc gcgcctgagg gagctcgccg cgcggcgagc 3360
gccctccacg agggccgggc cacctcggac agcgtctcta ctcctccgag gccgccgcgc 3420
ccccgcccc ggccgccgcc gccgccgccg gctccagctc gcagcgcacc accaggacct 3480
cgccatccgc gagctccggc cgctccacga gcgcgtgcgc gcccgcgcgc accgccgtga 3540
gcacgtctcc cagcgccggc ttcagccgcg ccagcgtcgc ggcgttcgcc ccgagcgcga 3600
ggtcggtcac gacgcgcccc acgctcgcgc cgagctcgct cttgcgcttg ttgaccgccg 3660
ccatcgccgc cgccgccaga tccaggagcc ccggatccga aggcgccgcg accgccgcga 3720
aatccgccgc tgaaggccac ttcgcccggt ggatcgaggt atcgcccgtc tcctccgcgt 3780
acacccagcg ccagacctcg tcggtgatgt acggcaggac cggcgcgaac agccgcagca 3840
gcaccgacag cccgagccgc agcgccgcca ccgccgagcc gcgcgccgcc tccccggcgc 3900
cgccctcgcc gcgcgcccgc gccttcgcga gctccaggta ggcgtcggtg aaccagcgcc 3960
agaagaaatc ctcggtccgc tcgagcgccg ccgcgaactc gtgctcgtcg aacgagcgcg 4020
tcgcgtcgtc caccacggcc gacagcttgt gcaagagcgc ccggtcgagc tcctcggaga 4080
tcgggtggac ctccgccgac tggctgagca cgtacttgct cgcgttccag atcttcgtga 4140
cgagccgctt gccgatcttc agcaccttct cgtcgaacgc cgtgtccgtg ccgagccgcg 4200
cgctcgccga ccagtagcgg accgcgtccg aagaatacgt gtcgagcagg tgcatcggcg 4260
tgacgacgtt gcccttgctc ttcgacatct tcttgcgatc cggatcgagg atccacccgg 4320
agatcgcgac gtggtgccac gggaccgacg actcgtgcag catcgccttc gcgatcgtgt 4380
agaacgccca cgtcctgatg atgtcgtggg cctgcggccg cagatcggcc gggaagagcc 4440

gcgcgtggcg cgccggatcg tccccccagt gagagctgat ctgcggcgtg agcgagctcg 4500
tgaaccacgt gtcgaagacg tcggactcgg cggtgaagcc gccgggctgg tcccgctgcg 4560
acgcctcgta cccgggcggc acgtcgaccg tcgggtcgac cgggagcatc tcgcgcgtcg 4620
cgagcagcgg ccggctgtga tccgggttgc cctcggcgtc gagcggatac cagaccggga 4680
actgcacgcc gaaataccgc tggcggctga tgcaccagtc accctggagc ccctcggtcc 4740
agttgcggta ccggaggcgc atgaaatccg ggtgccactt gatcttgtcg ccgtattcga 4800
ggagctcggc cttcttgtcg gcgagccgga cgaaccactg ccgcgtgggc acgaactcga 4860
gcggctggtc gccccgctcg tagaacttca ccgcgcgctc gatcggcctc ggctcgcccc 4920
gcagcgccgg ccccccggcg ggcgccgccg cgtgctcctc gcggcggagc agctcgacca 4980
ccgccgcgcg cgcctgcttc accccccctgc cctggagcgg cgcatacgcg gcgttggccg 5040
cggccgggtc gcggctctcc cacgcgccct cgccgaacgt caccggcagg acacggccgt 5100
tcttgccgag catctgccgg agcgggagct tctgctcccg ccaccagatc acgtcggtcg 5160
cgtcgccgaa ggtacagacc atcaggatgc ccgtgccctt ctcgcgatcc acgagcgggc 5220
tcgggaagat cggcaccggc gcgcggaaga tcggggtgag cgccgtcttg ccgaagaggt 5280
gctgataccg cgggtcctcc gggtgcgccg tgacgccgac gcaggccgcg agcagctccg 5340
ggcgcgtcgt ggcgatgacg agctcctcgg ccgtcccctc caccgcgaac gcgatgtcgt 5400
ggaacgcgcc cgattgcggg cgatcctcga cctccgcctg ggcgaccgcg gtctggaaat 5460
cgacgtccca catcgtcggc gcgaagaccg agtagaggtg gcccttctcg tggagatcca 5520
ggaacgacag ctgcgccgtc ctgcggcagt gatcatcgat ggtggcgtac tcgttccgcc 5580
agtcgaccga gaggcccacc cggcggaaga gcgccttgaa gacctgctcg tcctcgcgcg 5640
tgaccttgtg gcagagctcg atgaagttgg gccgcgacac gatgcgcggc ggctccttct 5700
tgatcgtctc cggcgcggcc tgcggcaagg tcaggccgcg ctcgtacggc gtgcgcacgt 5760
cggtgcggac gtggaagtag ttctgcacgc gccgctcggt gggcaggccg ttgtcgtccc 5820
agcccatcgg gtagaagatg ttgaagccgc gcatccggcg ctggcggacg acgacgtccg 5880
tgtgcgtgta gctgaagacg tggccgatgt gcagcgagcc cgaggcggtc ggcggcgggg 5940
tgtcgacgac gaaggtctcc tcgcgggggc gcgacgggtc gtatcggtac gtcccgtcgg 6000
cctcccacag gtcggccagg cgcagctcgg cgacgggcga gtcgaagtgc ttcgggagcg 6060
tcgcgggatc gatggagcgg aacgtcttct tgatcgtcac gtggtcacct gcagaacaga 6120
ccccgcagga accgccccgcg gggccggcat cctacgtcgt ccccccgggtg ccgctcaagg 6180
cgcgccgcgc ccgcgcggcg gcgatccgcg atcgcatccg cgcatccgcc agagccccggc 6240
ggctccgccg gcgcgcgcgc gccgtccgtg gagccgagag gagaggccgg cgcccaggtc 6300
gtggaggacg ccggcggcgc cgccgcggag atcgcggaga ggcgggcgca tcgatcgcgg 6360
cgaggccggg ggctcagtcg tagcgctcga cgtggacgtg cttgcggtgg acgccgagct 6420
cgccgcgggc gagctcgcgg acggacgaga ccatccgatc caggccgcag atgaagacgt 6480
gcggcgccgg atctccgctc ttctccgcga gctcccggta gagctcgggc acgtgcgcct 6540
gcacgtagcc gcggcggccg gcccacgacg ggccgccgcg cgagagcgtg atctcgtagc 6600
ggatccggtc ggatccgcgc gcgagcgcct cgagctcgtc gcggtagatg acgtcctcct 6660
cgaagcgcgc gccgaacagg atccacaggt ggggcgcggc cagccccgcg cgcagggagg 6720
cgcgcagcat gctccggagc ggcgtgatgc cggtgccggt cgcgacgaac aaggagggcg 6780
cggaatcccc gggatcgcgg gtgaagagcc cgtgcgggcc gatggcgcgg agcgtggcgc 6840
cgggctcgag ccggtgcagg tgctccgagc ccgccccgcc ctgcacgagc gtgaccgcga 6900
gatcgaagcg gggcgagccg tcgggcgcgg atgcgatgga gtaggcgcgc ttcacctcgc 6960
cgcccgggag cgggaggacg aggttgaccc actggcccgc ctcgaacaga aacgacctcc 7020
cgtcggcgcg ctcgaacgag agctcgcgca cgaaagggct gaggggccgg gcggcgacga 7080
ggcgggcttc gaacggttcg gcgtggatca tggtcggggc ccggcggggc tcggctgcga 7140
ggccgcgcgg gtggcgaggt cttaccgcag cctgcgcccc ggcccaatcg cgatcgccgc 7200
gggaagggcg ccgccggagg gcgcgcaatc gcgggaatca cgggcttccg ccccgtgcgc 7260
cgccggagcg cgcggccggg ccgccggccc gcgctccggc ggggagccgt cgcgggctct 7320

accgcacgcc catgcggcgg cgctgcggga tgttcaccgc cggccgggag cgatcctggt 7380
tggggagcgc gcgcggcggg cggggatccc ggtgcgcggg cttctgcgcg gggagctgcc 7440
ctcgctgagc cgggcgctgg tcgggcgact tggccgagcc cagcgcgaga tcggagacgg 7500
ggagatgcgc gcgtcgctgc atagaatcct ccatggaatc ggtcatcaac acatcgggaa 7560
gagcacccag gctgaaagaa accttcgaag aaccggctct catacaccct ccattcatcg 7620
tgcgaccccg gattcaggac ggatcgaacc cgcgagggac gctggctctc tgggcctctc 7680
cctgctcgct cgaccggcgc cctctcgacg caactccgcc gttcgtcggg acgggacggt 7740
ccgcctcgcc gcacgctccc cgtcgagacg actcagcgtc tcgacgtcag gagagatgac 7800
gactcggccc gtcgcgccac gacccttccg gctcggtgct tcgagcgcgc ggccagcgag 7860
cgaggggcga tcgccaggag atcacgaatc tcccggccat cggcctccag cgcctcgggc 7920
tcgttcgctc gtcgccccgc tccgtccccg cgcgcgcacg acgcgagctc gcgcggggaa 7980
ccgcgggccg ctgtcgtggc tgctgatgcg cgacgataca ggggggacgc cgtgcctacc 8040
tgggcaacag gcgctcatct tctaccacgg cgagcactac ggtgagtgct gccatgagta 8100
ggcccctgag ggtccgcgcg acggagcgtg gtgtcagcga gagatgcgca tggtggacgc 8160
gggctacgcg tcgagaggga cactagcact cgacctcgat cctgcccagc acttttttgtc 8220
ggggagggct gccctcccgc tggccgctgg ccgctggccg ctcgccgctg gccgctcgcc 8280
gctggccgct ggccgctcgc cgctggccgc tggccatgtg cgacgtgagc tcgagcagcc 8340
cgcggctgac ggacagaccc cggagttcat cgagccggtg atgccgaacc cgccaagcga 8400
aaaaacgtat ccgttcggca ggtcgtggcc tatcatgcaa gctgctcgat gcgctgacag 8460
gcttcttcga gatcctcgtc ggtctttgcg aagcaaaacc gcatgaagcg actcccctgc 8520
gtcccttcaa agaaggcgtc gcctggcacg cccgccaccc cggtctcgtc gagcaagtag 8580
atggctcgct ctcgacctgt cctcccgggt aggcgagaca catccgccag cacgtagtac 8640
gtcccctgcg gcacgcaggg tggcaagccc gctttctcca gcgcccgaca gaaccggtct 8700
cgcttccgtt catatcccctg ggcaagcccc gtgtaaaacg agcgaggaag gccgcggatc 8760
ccggcagcga ctccatgctg cagcggcgtc ggcgcgcaga catacagcag gtcgctcatg 8820
gctccaatgg ccttcgccca cctggcatcg gccacgctgt agccgatccg ccatcctgtg 8880
atgctgaagg tcttcgagta gccgcctatc gtgatcgtac gctcggacat gcgcggaagg 8940
gaggcgacgc tgacgtgctc acggccgtcg aagataaagt actcgtaaat ttcgtccgtg 9000
atcaccatga ggtcatggtg gcaggcgaga tcggcgatct gttccagctc cattcggccg 9060
aacaccttcc cggaaggatt tccaggagag ttcaccacga tcgccttggt cttcggggtg 9120
atcgcgcgct ccagctcgtc gccgtcgaca ttccagctca gggatcgcgc cgtcacatac 9180
cgcggaacag cctcgacggc gaggatagcc tgggcgtgat aggcataaaa cggctcgaag 9240
agcagcactt cgtccccagg attgagcaag gccatgcaag tggcctgaaa ggcccctgtc 9300
gctccggcgc tcaccgtgat gtcagtctcc ggatccgccg cgatgccatt atggcgagcc 9360
agcttcgccg cgatcgcatg gcgcagctcc acgatgccgt cgaagcgcga atatgtattg 9420
caccccgat ccatcgcctc cttcaccgct tgaaggatca ccgaaggaac tggggtatca 9480
cagacgccct gggacatatt gatcccatgg accttggcgc acgccagggt catggtacgg 9540
atatcggact gggcgaggcg agccgcacga tcactcggta gactcttcat cagcgtgctc 9600
ctgcttctgt tctgcggctc tgcatggtgt cttcgggtgg gcttgtcagc tcgacgcgcc 9660
catgcagcgg cgcagcccta gcggccgcag gtctgtccac acttctttga tgaaagcgag 9720
acattcggct ttcgtgccct gtttgcccgc agccctccag ccccaggta cgggcttgtc 9780
ggcgggccag atcgagtact gctcttcgcc gttcaccacg acctggcaac gcgtcttgct 9840
ttcgtcgtcc cgattcatga ttttcctcgc ccttcgtcag cgctgcgcga gcatgaaacg 9900
aatcgctcat cggcgcacag gcgcgcgccg gctgcccgga ggcactccca cgcctccctc 9960
acggcaacct catcgctccg gatgttcccg atggcgactc ggatcgtgta cctgccgtgg 10020
agacgggtat gggacaaaaa taccctgccc gacttgttga cctcgtccag cagcgcctcg 10080
ttgaggcgat cgagctcgcg ttcgatcgac tctctctctg cctcgtccgc cgaccgcatg 10140
atgcaagcga gcgcggaggg cctcatgcga aagcagaccg tactgaacgg cgtcggcgcg 10200

aggcgctccc aatcgggatc ggcgtccacc cactgggcga gctgctgccc caatcggagg 10260
tgctcccgga tccgggccgc cagcccttca tgcccgaagt agcgcacgat catccagagc 10320
ttcagcgctc ggaagcgccg accgagctgg ataccccagt ccatgtaatt cgtgacgtcg 10380
ccctcggtgc ggaggtattc gggcaccaga ctgaacgcgc gcttcagtcg gtcggcgtca 10440
cgcacgtaga gcacgctgca atccatgggg gtgaacagcc acttgtgagg gttcactacc 10500
agcgagtccg ccccctcgca gcccgcgagc acgtccctgt gctcggggac gatcgcggcc 10560
atccccgcgt aggccgcgtc cacgtgaagc catagcccgt gctcccggca aacgctgacg 10620
atggcgggga tggggtcgac gctcgtcgtg gacgtcgtgc ccaccgtcgc cgcgacgcag 10680
aagggtcgga ggccggcccc gaggtcctcc acgacggcgg cgcgcagcgc ctcggggacc 10740
atgcggaagg ccggatccgt ggggatcttc cgcaccccct cctgcccgat gccgagggtg 10800
atggctgcct tctcgatgga tgagtgcgcc tgctccgacg cgtagagtcg catgcgccgc 10860
tgtcccgcca tgccccggag ccggatggtc ggctcggccg agtcgcgcgc ggccgcgatc 10920
gcgaccatgc tggcggtcga cgcggtgtcc atgatcgcgc cgtgcaagcc ggcgtcgaga 10980
tccagcatct gacgcagcca ggagaggacg agctcctcga gctcggtggc cgccggcgac 11040
gtgcgccata gcatcacgtt gacgttgagg cacgccgcga gcagctcgcc gaggatccca 11100
ggaccagacg ccgtgttcgc gaaatacgcg aagaatcgcg gatgattcca gtgcgtgatc 11160
cccggcagaa tgatctgctc gaaatcggtg agcacggcgt ccatcggctc cggctcgacg 11220
ggcggggtcg gggccagcct gcccttcacg tcgccggggc ggatcgcggg aaagacgggg 11280
tatcgatccg ggtggccgag gtaatcggcc gcccaatcga tgattctcat accgatccgg 11340
cggaactcct ccagatccat gtcccccgagc cgttctttcc gcgggtcgct cacgtcaacc 11400
tcctcgccct gccaggacag gatcctcgag gtcccctggc tccggcggtg gaaagcgctc 11460
cttgaacgtg aaggcccacg gggtcggtcc gtagcgccgc aggtgctcga gccgatcctg 11520
cccctcgcgg acggacggga tgtgcccggc cgggacccac cacagcacga ggtaatgcgg 11580
ctcgagatgc tcgaaccacc gagcgcgctg tcgcaggaac gcggcatgat ccgcggtgta 11640
ggtgaaggcg aacaggtgct cgatggaggt ccataccgac agggtcacga ggagccgctg 11700
gtccgggtac ggacggatgg acacagagtt cccctcggcc gtctgcaggc gccacacgaa 11760
cccctcgctc cgatcggcca gatggttgat atggtcgagc ccctggacga agccctccat 11820
gatcggatcc tccagcggag cgcgaataca tgcgaagttg tattgcgcga tgtggtgccg 11880
atgctccgac atgtcgcttt ccatctccag ctcccgctca ccaatcccag cgctgctccg 11940
gggagctcat cagggcagac gcgacatcga tcccgaagct ccgccgcatc ccctcgacga 12000
aggcggcctg gaccgcttcg gcgacggatc ggcctgcctc cggcaagacc tcggagacaa 12060
agaagaaccg cctcgtggaa gggacaatct tgcccccgctc cgcctggcgc catacgaagt 12120
gcctcgtcac cagtcccctcc gcgtcggcat acccgacctc gccggcgccg accgccacgc 12180
tgccgcctga gccgagctcc acgaacgcct caccgcctcg cgagatctcg aggcgaacgt 12240
ccgggccagc cagatcgccg aggtcccaag cgccgacggg gacggcgaac cgcagcgaca 12300
ggaggttgta aaaatcgacg aatgcgttga tgtgcggcag ctctccaacca ccgaggaccc 12360
gcttcgccag cgcctcgatc gagctcggaa atttcttgcc agagacccccc actcgcttca 12420
tcgcctcgcg ccaggcagcc acgtgcggat gcgactgggc gttttcgtgg ccccagctcc 12480
gtcgcagctc ctcctcgacc ttccggagct cctccagcac ggccggccgc tctgcggcgt 12540
tgtccaggcc ttccccgtac ccggtgacca agatcatccc aggaaacgac tcccagattc 12600
gcggatcgac gatgaatgcc atgtgcctcc tgcccctcga gagcgatcgc ctcgatcgac 12660
accaggctgt ggatgcatga gccgggccgt gcggacgcag gaccccgcta ctcatggctc 12720
ttcgtggccg atgaacaggt cctccacccg tcgatcgtgc tcggtgcccc gatccgtcca 12780
gtcccacccg ccggcgaccg cgatgtttgc acccgagacg tacgaggcgc ggtcggaggc 12840
gaggaacgcc acagcatctg cgacctcgct ggcgcgcccc aggcggccca tggggacgcg 12900
ccgctccatc cactccttct gcgcgggcgg aaggtatccg ttgtcgatga gccctggaga 12960
cacacagttg accaggattc catgaggcgc ctcctccgtg gccaggctgc gcgtgaggat 13020
gagcacgccg gtcttcgcga tcgagtacgc cgccacgttc ggcgcgccgc ggatcgcgta 13080

cgtggggctc aacccgatat tgatgatccg gccgctcttt cgctggcgca tgcgcgccac 13140
ggccgcgcga cagaggtaat gaacgctgct caggttgctg tccatgacgt tgcgccattc 13200
gtcgtcggtc atcgccgcaa gcggcttgaa gaagaagtcg cccacgttat tgacgaggat 13260
gtcgatgggg cccagctgcg cctcgacgct ggagaagagc tccgcggccg cgttggggcg 13320
ggtgacgtcg gcctgcacca ccatggttcg tcgcccgagc gcgcggatct cggccgccgt 13380
ctgctcggcc gcatccttgt tcgaatggta attgacggcg acgtccgcgc cttgctccgc 13440
gaggcgcagc gcgatcgcct tgccaattcc gcgcgagcta ccggtgacca gggcgacgcg 13500
cccggcgagc tccagcgatc gcgcctgtgg cagggccgga gcagcctcct ggtggagctc 13560
gacgtcgacg gggagctcca cgtggtagct cgtctctcgc ggagccgcgc agtacctctc 13620
gtagaacgcc tcgaggacgg gctcttcgcg tcgcatgatg tccgcgtggg attcggcgct 13680
gcgccacgga tagagcacca ggatctcgtc ggggcgcacg gtgctctgaa agaatcgcgc 13740
atgaccgcac caccccgggt gttcgtgggc tcctgggccg agcagatcat ccatttttttg 13800
catgatctgc gtggcctcgc cctccatgcc gggtttgatg cgccatcgct ccataacgag 13860
gatcatgtct tggctcctgt tcgtcatcgc cgtttcgatc tgggggggct gcccgcgctc 13920
tcgagggcgc gccccttgta ttggccgcgg atggtctggg tagcgctcgc gagctttcgc 13980
ttgtgggcgg cgttcaggct tgcgccttga ttgacgaacc gctcgcagac gaatgcgtgc 14040
gattcatatg cggtcgcgag cgcccacagg tagagacggt ggctacgagc gagcttgggc 14100
ggaagcgccc tcaaggcggg tcggatggat ttgagggtgc gaatcaagcg ggcgtgctct 14160
tcgaagaaga gaccgctgaa tccatcctgg acgaacggcg gggccatggt aggtcgcacg 14220
atgttgttgt aatcatcatt cgagaaatcg ccggtaaacc ggaaagatgc cgccgttgcc 14280
caccagagct gcgtggagat cgcgagcgcc tctcgcgccc tcggaacgtc acggttggcg 14340
agcgcctccg acagccacct gtgcgccatg agcagcgcct ggaccagcac gaagaacgcg 14400
tgatgcccca atacccagcg gcccagcgcg ccgtccggag cgcccgcctc cgccggaggc 14460
gcagccagct gcgggatgcc attccatgat tttggccttc gcttgccgga gaactggtgg 14520
aggatgtcct cgatggacgc acagagatgc cccatttcca tgggttgcag ggaagtacct 14580
ttcaggcttt cgcggatcat tcggtaatat gcgacgatca ccgcttcgca gtacgtctcg 14640
agcgggtctg gggcggagac ccggtgcaca tggaaatagg cgtcgtactc cgcctcccgg 14700
tccgacaggc tccccgggat gtccggatct gccgcttgcg cctgccatcg ctcgaggatg 14760
gggacggcga ggggggcgag atgctcggcg agcaccgcca ggtcgccctc cggcgcatgc 14820
gccaacaaga cctcgaaggc ctcggctgcg accttggagg tctcgcccat tccgacggcc 14880
tcgatggctt ggggcagcgg tagatggatg gtatatttag ccatgatttg cccgaagatt 14940
gccgctgcgt cgacagatct ttcgcgagcc ggaacgccat ttccactgct ctggctctca 15000
atattgaatt gagccctggc gactgccata ggcccagtcg ctcgacacag tgtacggagc 15060
ggcccgatgc tttctccttt tttagtcctg caccgaatac ttctgttggg cgccaaagat 15120
cccttgccga gactgtccgg cgagatgtcg tgtgcgaagc gtccgcacgt ccagcgggcc 15180
catgcgttgc tagagcataa aacggttcga tgcctggtcg agagggagac gcgaggagcc 15240
tccctttggg acggatgagg aatttcgtga ccgaaatgtc ggcaggaaca gcggcgcaga 15300
agcggcgcat cgatggggaa ccatgggtta cgaagacatt gatgataatg tcgacgcaat 15360
cgcaatcgtc gcgatgagcg gccgcttccc cggcgcgaga aacgtcgagg agctgtggca 15420
gaagctccgc gctggcgtgg aatgcgtcgt cacccttcaca gaggccgagg cgctcgccgc 15480
gggggtgagc cgcgagatgc tcgcgaatcc cagctacgtg cgcagaggcg cgccgctcga 15540
cggcgtggag ctcttcgacg cctcgttctt cgggttcagc ccgagggagg cagagagcat 15600
ggatccgcag cagcgcatct tcctggaggt cgcctgggag gccctcgagc gcgccggtta 15660
cgaccccgat gcccattccg ggcctatcgg cgtcttcgcg ggcagcgccc cgagcggcta 15720
ccactccctg gcgcagtccg acccggagat cctaggcgcc ctcggccact accaactgac 15780
gctgaacaac gacaaggatt atctcaccac acacgcctcg tacaagctca atctgcgggg 15840
cccgagcgtg tgcgtgcaga cgtcctgctc gacctcgctc gtggccgtgg tcatggcctg 15900
ccagagcctg ctcaaccacg agtgcgacat ggcgctcgcg ggtggcgtgg ggatccatgc 15960

gcatcagcgg aggggctatc tgtatcagga gaacggcatc tcttcgcccg atgggcattg 16020
ccgcgccttc gatgtggccg ccaagggcac cgtgggcggc agtggcatag gcatcgtcgt 16080
cctgaagcgg ctcgccgacg cgctcgccga cggcgaccac gtgcacgcgg tgattcgagg 16140
agcggcgatc aacaacgacg gctcgagcaa gatcggttac accgcgccga gcgtgcaggg 16200
gcaggccgag gtgatcggca tggcccaggc gctcgccggc gtggagccgg atgacatcag 16260
ctacatcgag gcgcacggca cggggacgcc gctcggcgat cccatcgaga tcgcagccct 16320
cacgcgcgtg ttccgggcga agaccgcacg aaggcagttc tgcgccatcg gctcgctcaa 16380
gaccaacctc ggccacctcg atgccgccgc gggcgtcgcc tcgctgatca aaacggtcat 16440
ggccctcgag caccgcgagc tgccccccgag cctgcacttc gagcgtccga atccgaagct 16500
cgagctggag agcagccctt tctacgtcaa cacccgcctc actccgtggc acgcggcacg 16560
aggtccgcgc cgcgctggcg tcagctcgtt cggcatcggc ggcaccaacg cgcacgtggt 16620
cctcgaagaa gctccggccc cgcctccgag cggcccctcg cggcgttggc agctcctcac 16680
cctcgcggct cgctccgagg ccgggctcgc gcgggccacg gccgacatga tcgagcacct 16740
cgatcgccac tccggcacat cgatcgccga tgtcacgtac acgagccacg tggggcgccg 16800
ggcctggccc ttccggcgag cggtcgtcgg cgagagcgcc gcggatctcc gcgccgcgct 16860
cgcgagcgag ggctcgccgc gctcgatctc gtcatgccag gcggcgaggg agaggcccgt 16920
cgtcttcctg ttccccggtc agggagcgca gcacctcttc atggcgcggg agctgtacga 16980
ggtcgagccg atcttccggc agtccctcga ccgctgcgcc gagctcctgc gcggcccgct 17040
cggcctcgat ctgcggcagg tcctctaccc cgccgagggg cagcgcgacg acgccgagca 17100
ggagctcggt aggaccgcga tcgcccagcc cgcgctgttc gccatcgagc tctcgctcgc 17160
caagctgtgg atggcctggg ggatcgtccc ccaggcgatg atcggccaca gcgtcggcga 17220
gttcgccgcg gcttgtctgg cgggcatctt ccgcgaagag gacgcgctcc gcctcgtcgc 17280
cgagcggggc cgcctgatgc aacagatgcc gcccggcgcg atgctggcgg tgccctcgc 17340
ggagcccgag ctcgcccctt acctcagcga cgacatctcg ctcgcggcga tcaacggtcc 17400
ggctctctcg gtggtcgctg ggccgatcga ggccatcgac gcgctcgcgg ccgagctctt 17460
ggaccacggg ctctcgtgcc ggcgactcca cacgcggcac gccttccact cgaagatgat 17520
ggcccccgtc gttgacgcct ttacccgatg cgtgtccgcg gtcgagcgcc gcccgccgtc 17580
aggccacttc ctctcgaccc tgacgggcgg ctggatctcc cccgaagcag cgaccatccc 17640
cgcatactgg gcccggcagc tcgtggagcc ggtgcgcttc gcccaggccg tgaggcagct 17700
gctgtccgag tcgacgtggc tctggctcga gctgggtccg ggccagaccc tgagcccgct 17760
cgtacggcag caggcccgcg cggatggcgg ccaggtggtc gtcgcctcgc tgccgcgcgc 17820
gaaggacgcg ggcgccgacc acctcgcggt catcgaggcg ctcggccgtg tctggagcgc 17880
tggtgggacg gtcgactgga agcgctttca cgagggcgag gcgcggcggc gggtgctgct 17940
accgacctac cccttcgagc ggcaacgata ctgggcctct ccgcgccaca cgagcgctcc 18000
gccggaagcg ataatcaagc cgctcctcgc gaagaaccca aacgtcgccg attggttctt 18060
cctccctgcc tggcggcgct cggatcctcc ggtctcgttc gacgcgcagg cggtgaccac 18120
gcggcgctct acgtggctcg tcttcatcgg ggacgagggc ctcggcgcgg cgctggtgga 18180
gggcctcgcg cggcgggggc acgaggtcgt cgcggtggtc acgggtgaga ggttcgagca 18240
gacgggcacg cagcgctaca cgatcgatcc cgccgcgaat ggcgatgttg cgtcctcttt 18300
cgcgcggctc gaaatcgaag ggcgcatgcc ggaccggatc gtccatgcct tctgcacgtc 18360
gcctgcggac ggcgcgcgca tcgagcgcgg agccgcgctg gagatcgagc gcaggctggg 18420
cttcgatagc ctcctcctcc tcgcccaggt gatcgccgca caaaggcatc cgaagccgct 18480
gatgctcggc gtgatcacga cccgggcgca ctccgtcatc ggaaccgaga tcatcgagcc 18540
cctgcgcgct ctggtgctcg gcccctgccg cgtcatcccg caagaaatac cccatgtctc 18600
gtgccggaac atcgatatcg atctcccggg cgaaggcggg cgcgcggaga tcgcggcgcg 18660
cctgatcgcc gatctggagc gagagtcgcc cgactcggtg gtggcctacc gcggcggccg 18720
gcgctgggtc gagagcatag agctcaccga tgtcggccgg cggtcagctg gcgccgcccc 18780
gcgcctccgc cagcgcgggg cgtacctcat taccggcggc ctgggggggca tcggcctcgt 18840

ggctgcagag ctcttggccc gagaggcgca cgcacggctg atcctggttg ggcggacagg 18900
cctgccagcg cggcaggggt gggacgactg gctcgcggcg cacggcgcgg gcgacgcgac 18960
gagccgaaag atcctccgga tccgcgcgct cgaggaggcc ggcgccgagg tgaagatcgc 19020
cgcggccgac gtctccgatt tcaatgcgat gcggagcgtc atcgaggagg cccggacgcg 19080
cttcggccgc atcgacggcg tcattcactc cgccggcatc gcgagtggag gcatgatcca 19140
gctcaggacg ccgatggcgg cttggcgcgt gatggcgccg aaggtcggcg gcacgctcgt 19200
gctcgatgcg ctcctccggg acgagcgtcc cgacttcctc ctgatctgct cgtcgttggc 19260
ctcgctggtc ggcggcgcca cccagatcga ttactgcgcc gccaacgcct tcctcgacgc 19320
ctacgcgcag agccgcgagg gcgaggaggg atgccgcgtc atctcggtgc aatgggacac 19380
gtggagtgac gtcgggatgg cggtggactt caagctcccg gccgatctcc aagaggggcg 19440
ccgcgagagc ctgaagcggg gcatcagctc gagcgagggc gccgaggtgc tcggccgcat 19500
cttgagcgca ggcatgagcg gcccgctggc gatttgcacg tcggatctac cagcgtacaa 19560
gcagtctgtc acgacacgcc gatcgcagca tgagcaaact cccgccgccc ggccgatgca 19620
ctcgcgccca acgaccacgg gagcctatgt cgctcccgag accgagaccg aacggcgcat 19680
cgccgcgatc tggcaggatc tcctcggcct cgagcaggta ggcgcaaacg acgatttcct 19740
ccagctgggc ggccattcgc tgttggccac gcaggtcctg tctcgcgtcc tgcagaccct 19800
caaggtgggg atctcgttgc cgcagttctt cgatgcgccg acggtcgcag ggctttcgcg 19860
cctggtcgac gcagcacggg ccgaaggcgc cggacccgtc gcgccggcaa tcggccgtgt 19920
cgagcgagac gcctaccgaa tcaagccgcc cgcggccgaa caggccgccc gcaccaagcc 19980
gtaacaagaa ggggatcgag tcatggaacc cgtcggcggc gtggacatga atcagcccgc 20040
aaagcagcag gagacctgcg tcttcccgac ctccttcgcg cagcggcggc tctggttcct 20100
cgaccagctc gagccgggga gcgccgtcta caacatgccc gcctccttcc ggacgcgcgg 20160
gccgtacgac gtcgactcgc tcgtgcgcag cgtgaacgag atcgtgcggc gccacgagtc 20220
gctgcgcacg accgtcgatg tcatcgatgg cgaacccgtg caggtgatcg ccccctcgct 20280
gcgcatcgag gtgcccgtcg tggacctgag cgagatcgac gagccggagc gagaggcgga 20340
ggcccggcgg ctcatggcgg aggagagccg ccgccccttc gatctcacgc gagggccgct 20400
gctccgagcc aagctgctcc ggctcggcga ggccgatcac gtgctgatct tgacgatgca 20460
tcatatcgtc tccgacggct ggtcgatgga cgtgctgttc aaggagcttt ccacgctcta 20520
cgccgccttc cacgagggcc gcccgtcgcc gctcccggag ctgccgattc aatacgccga 20580
cttcgcggtg tggcagcggg agctgctcca gggcgaagtt ctggaatcgc acctcgggta 20640
ctggagagag cacctccgcg gcgccccccac gctgctggag cttccgatgg accggccccg 20700
gccgccggcg cagacgttcc ggggctccca gcgcgcgttc cgactcccac tctccctgca 20760
acaggcggtg caggcgctca gccggcagga aggcgcgacc ccttcatga cgctgctgac 20820
ggcgttcagc gtgctgctct cgcgttatgc gcggcagagc gatctggtgg ttggcacgcc 20880
catcgcgaat cgcacccgag cagagctgga ggggctgatc ggcttcttcg tcaacatgct 20940
ggcgctgcgc atcgacctcg ggggcgaccc gagcttccgc gagctgctcg ggcgggtgcg 21000
ggaggtgacg ttgggcgcct acgcgcacca ggacctgccc ttcgaacggc tggtggagga 21060
gctgtcacca gggcggagcc ccagccacag ccccttgttc caggtgtcct tcacgttgca 21120
gaacaccccg atggatgcga cgaacagagc agacattgca tcgggtggcg cgccgctggt 21180
ggaaatgaag gcggcgaaat tcgatctgat cctggagctc tcggaatcgc cgcaagggtt 21240
gctcggcacg ttcgagtaca acaccgacct gttcgacgcc ggcaccatcg agcggatggc 21300
cggccacctg gaggtgctgc tctccagcgc cgtcgcggcg ccggatcgac ccattgcgga 21360
gctgccgctc atgggggccg aggagcgcag tcgggtattg gtggagtgga actccactgc 21420
cgcgctgtat cccgaggacc attgcatgca cgagctgttc gagcagcaag tggagcggtc 21480
gcccgaggcg accgcggtgc tcctccagca gcagacgttg acgtatcgag agctgaacat 21540
gcgcgccaat cagctcgcgc atcacctgcg gagcctgggc gtgggcccag aggtgcgcgt 21600
cgggttgtat ctcgaacggt caatcgagac ggtcgtggcg atcctcggcg tgctcaaggc 21660
tggcggggcc tacgtgccgc tcgatccgac gtaccccagc gagcgcctcg ggctcatgat 21720

ggcggacgca gcgccctcgg tgctgctcac gcaggcgtcg ctcctctcga agctgccgcc 21780
ccacggggat gcaacgctgg tacagctcga cgcgctgcac gaagcgctct ccaggctgcc 21840
acaccatacc ccgcggagcg gcgtcaccgc ccagaacctc gcatacgtca tgtacacttc 21900
cggctcgacc gggcggccca agggcgtgct cgtcgagcac cgcggcctct gcaacctgcc 21960
caccgtgcag gccaagctct atggaatcgc gccgggcgac aggctcctcc agttcgcgcc 22020
gctctgcttc gacacatcgt tctgcgagat cgcgctcgcg ttgctctcgg gagcgacgct 22080
ggtcatgggc acggcggacg agcttctccc gggacctccg ctggtcgagc tgctgaagaa 22140
gcacgcggtc acggcgatgc tcctggcccc taccgtgctc gcagcgctgc cagaacaaca 22200
gagcgcggcg ttgccgctgc gcgtgctcac gatggccggt gaggcgtgcc cggcggagct 22260
cgtcaagcgc tggaaggcac ccggacggcg cctgttcaac tcctatggcc cgaccgagac 22320
gaccatttgg gcaagctccg cagcggacct gtccgacgaa cggatcccgc ccatcggccg 22380
tccgattgcc aatacgcaaa tctacgtgct cgacgaagcg ctcgagccgg tgcccatcgg 22440
cgtgccgggc gagatcttca tcggcggcgt gggcgtcgcc cggggatatc acgggcgtcc 22500
ggacctgacg gccgagcgat tcgtacccga cccccttcggg caaaccaaag gggcgcgcct 22560
gtatcggacc ggcgatcggg cgcgctggct gccggacgga aacctcgagt ttctcggtcg 22620
aaacgacgag caggtgaagg tccgcggtgt ccgcatcgag ctggaggaga tccgcgcggc 22680
gttgctcaag cacccggcgg tcgctcaagc cgtggccgtg gtgcgcgagg acacgccggg 22740
ggacaagcgg ctcgtcgcgt atgtcgtcgg acgcggagga gcgcgcgtga ccgccgcgga 22800
gctgcgccag tccgtgagcg agcgattgcc tgcgaccatg gtgccatcgt ccttcgtggc 22860
gctcgacgcc ttgccccctga cgccgaatgg caaggtggac cgccgcgcgc tgccggagcc 22920
cgagcagagc gccggcggcg aggaccacgt cgcgccgcgc aacgccgtcg aggaggagct 22980
cgccaggatc tgggcgagcg tcctccggct cgaaagggtc ggcgtccacg acaacttctt 23040
cgagatcggc ggcgactcga tcctgagcat ccagatcgtg gtgcgcgcgc agcaggcagg 23100
gctgcgcctc accccgcgtc agatgttcca gcaccagacc atcgccgagc tttcgaccgt 23160
ggctagagcc gtcgaggcgg tccacgtcga gcaggacccg gtgaccggtc ccgcgccgct 23220
cacgccggtg cagcgctggt ggctggagca ggaggcggcc gagccgcacc acttcaacca 23280
gtcgatcttc ctcgaggtac gcgagcggct cgacgagagc gcgctggagc aggccatcgc 23340
gcatctgatc gaccaccacg acgcgctccg gttgcgcctc gcgcgcgacg aacgcggcgc 23400
ccaccaggtc ttcgccgcgc cgggaggctc gaccccattt cagcgcgtcg acctcggggc 23460
gctgcccagc gccgagcaga tctccgccat ggagaaggcc gcgagcgagg cgcaggcgag 23520
cctcgatctg gccgcgggcc cggtcgtccg cgccgtgctc ttcgacctcg gcgaggtcgc 23580
cccgcaacgg ctgctcgtca tcgcccacca tattgcggtc gacagcgtct cctggcggat 23640
cctgctcgac gatctctttg gggcctatga gcaggcgcgc cgcggcgagg ccgtacgcct 23700
gccgcccaag accacgtcgg tcaagcgctg ggccgagctg ctcaccgagc acgccggctc 23760
cgaggccgtc aaggcggagc tcggctactg gctcgactca tcgcgacgaa cggtagctcc 23820
gctgcccgtg gatcgacggg ccggcgagga cgtgtggggc tcggcgcgcc acatcgtcgt 23880
ctcgctcacg ccggagcaga cggagcagct cctgcgcgag gtgccgcagg cgtaccgcac 23940
acggatcgac gacgcgctcc tcactgcgtt cgcgcaggcc atcgctcggt ggacgggctc 24000
gccggcggtg ctcctcgacc tcgagggtca cgggcgcgag gagctcgccg gcgtagacct 24060
cacgcgcacg gtcggctggt ttacggccat gtacccgatc ctactccgcg tcgacgcggc 24120
ggatccgggt gaggcgctca aatcgatcaa ggagcagctc cgcgccgtgc caggccgcgg 24180
gctcggctac ggcttgttgc gttaccttcg gtccgatacc atcgccgagg tccgcgcgtt 24240
gccgcaggcc gagctctgct tcaactacct cggccagctc gatcaggcga tccccgaggc 24300
tgcaccgttc cggccggcgc gcgagtatca aggctcggag cgcagccccg gcgcccatcg 24360
cgcccacctc atcgaggtga acgcgagcat cgccaatggg cgcctgtacg ccacgtggac 24420
gtacagcgag cgccgccacg agcccgaaac catcgagcgc gtcgcggcga gcttcgtcac 24480
ggcgctccgc gcgctcatcg cgcactgcac cttgcccgag gtcggcggca acacgccttc 24540
cgacttcgac aaggtgcgcc tgcgccagga gaccatcgat gctctcgacg caatcgacgc 24600

gggccccggg ccgtctgcga gggggagccg aatcgaagac gtctacccgc tctcgccgct 24660
ccaggagggc atcctgttcc acacgctcta cgccaccgat tacacggcgt atgtcgagca 24720
gttccactgg acgctggagg gcgatttcga cgccgaggcg ttcacccgcg ccctccagga 24780
cgtcgtcgct cggcatgccg ccctgcgcac gtcgttcgcc tgggagcgcc tcgatgctcc 24840
acttcagatc gtccgcacgg gcgcggtcct ccccgtcgag caccaggacc tacgcggcct 24900
cgccgcggag gagcagaccg cgcacatctc ccgttacgtc gaggcagagc gccagcgccg 24960
gttcgatctg cgaaaggcgc ccctcatgcg cgccgggctg ctccggctcc gcaaggacgc 25020
ctggtgcctc gtcgagacca tccaccacct gatcctggac ggctggtcga cacaaatctt 25080
gctcaaagaa gtgttcacgc tctacgaggc gcaccgcgga caccgtgggc atctcgcgct 25140
ggagctcgag cagccgcggc cctacggcga ttacatcggc tggctcgcga agcaggacca 25200
ggtgcgcacc gcggccttct ggcggcgcga gctcgagggc ttctccgcgc cgacgccgct 25260
cggcgtcgac cgcgctgtgc cgcacgacga cggcggcccg cggtttggtt ggcgccgcat 25320
cgccctctcg ggcgacgacg cggccccggct cgccgccttc gcgcgtcagc atcagctcac 25380
gatgagcacg ctggtgcaag gcgcgtgggc gctgctcttg tcacgctaca gcggcgatcc 25440
cgacgtgctc ttcggtatga ccgtctcggg ccgctcggcg ccgattcccg gtatcgagcg 25500
catgaccggc ctcttcatca acaccattcc ggtgcgcgtg cgcgagcctg ccgacgcgtc 25560
ggtgctcgcg tggctcaagg cgctccagga gcacgaggca gagctgctcg agcacgagca 25620
cagcccgctg gtcgaggtcc aggcccatag cgacgtgccg cgcgggaccc cgctcttcga 25680
gagcctcgtc gtgttcgaga actacccggt gcaggtcatc ttcgaggccc ctccggtcga 25740
ggggccgacg cgcgcggagg agggcctccg catgatcgat gcgcagtata tcagtgatcc 25800
accgtatccg ctgacggtcg tcgcggcctt ccatgggacg ctttatctca atattggcta 25860
cgagcgccgc cggttcgacg accaggccgt cgaacggatg atcgggcacg tcacgacgct 25920
gctccggggc ttcgtgcaga ggcccgagac gtcggtccgc gatctgccgt tgctgacggc 25980
cgaggaggag cgcacccagc tccacgcgtg gaatgccacg gccgcgccgt atcccgaggg 26040
ccattgcatg cacgagctgt tcgagcagca agtggagcgg tcgcccgagg cgaccgcggt 26100
gctcctccag cagcagacgt tgacgtatcg agagctgaac atacgcgcca atcagctcgc 26160
gcatcacctg cggagcctcg gcgtgggccc agaagtgcgc gtgggcttgt gtctcgaacg 26220
gtcgatcgag acggtcgtgg cgatcctcgg cgtgctcaag gcaggcgggg tctacgtgcc 26280
gctcgacccg acgtacccca gcgagcgcct cgggctcatg atggaggacg cggcgccctc 26340
ggtgctgctc acgcagacgt cgctcctctc gaagctgccg ccccacgggg atgcaacgct 26400
ggtacagctc gacgcgctgc acgaagcgct ctccaggctg ccacaccata ccccgcggag 26460
cggcgtcacg gcccagaacc tcgcatacgt catgtacact tccggctcga ccgggcggcc 26520
caagggcgtg ctcgtcgagc accgcggcct gtgcaatctg cccaccgtgc aggccaagct 26580
ctatgcaatc gcgccgagcg accggctcct ccagttcgcg ccgctctgct tcgacacatc 26640
gttctgcgag atcgcgctcg cgttgctctc gggagcgacg ctggtgatgg gcacggcgga 26700
cgagctcctc ccgggacctc cgctggtcga gctgctgaaa aagcacgcgg tcacggcgat 26760
gctcctggcc ccttcggtgc tcgcagcgct gccagaacaa cagagcgcgg cgttgccgct 26820
gcgcgtgctc gcgatggccg gcgaggcgtg cccggcggag ctcgtcaagc gctggaaggc 26880
acccggacgg cgcctgttca actcctatgg cccgaccgag accaccattt gggcaagctc 26940
cgcagcggac ctgtccgacg aacggatccc gcccatcggc cgtccgattg ccaatacgca 27000
aatctacgtg ctcgacgaag cgctcgagcc ggtgcccatc ggcgtgccgg gcgagatctt 27060
catcggcggc gtgggcgtcg cccggggata tcacgggcgg ccggacctga cggccgagcg 27120
attcgtaccc gacccccttcg ggcaaaccaa aggggcgcgc ctgtatcgga ccggcgatcg 27180
ggcgcgctgg ctgccggacg gcaacctcga gtttctcggt cgaaacgacg agcaggtgaa 27240
ggtccgcggt atccgcatcg agctggagga gatccgcgcg gcgttgctga agcacccggc 27300
ggtcgctcaa gccgtggccg tggtgcgcga ggacgcgccg ggggacaagc ggctcgtcgc 27360
gtatgtcgtc ggacgcggag gagcgcgcct gaccgccgcg gagctgcgcc agtccgtgag 27420
cgagcgattg cccgcgacca tggtgccgtc gtccttcgtg gcgctcgacg ccctgccccct 27480

cacgccgaac ggcaaggtgg accgccgcgc gctgccggag cccgagcgga gcgccggcgg 27540
cgaggaccac gtcgcaccgc gcaacgccat cgaggaggag ctcacacgaa tctgggccga 27600
cgtacttggg gcaaagcggg tcggtgtgca cgacaatttc ttcgatctcg gcggccattc 27660
cctgctgctc gtccgggtgc atgatcggct cggccagcgg ttcgatcggc cgccctcgat 27720
ggtcgacctc ttcacctatc cgaccgtggc gtcgctcgcg cggttccttg gcgaacgggc 27780
gaacggcaag caatcgccga gggaggccgc ggcggacgtc acggagcgcg gccggcgccg 27840
cctggaggcg cgggcgcggc gggcgaaggc catccgtggc ccgacctgac ccgggcaccc 27900
ttccaagccc cgccgttcct cgcacatccg ccgcctcgag cgccgcgtcc agcgccgccg 27960
ttcgccgacg aggaggcgcg agacgacggt ccaaggcctt cgtgggctct ttgccccgca 28020
atccggaagc tgcgcggcag ttcgtcgccc ctgcaatgct gccattgtag agctcctccg 28080
ctcgccgcgg cctcttttct tgcggcccgt ccgcgattga cctcacatcc tgatcccttc 28140
ttgcgtcgtc cagaaagtga ttgacggcca gcgccgcgct tgagatcttc cggcgcgcgg 28200
cgatttcatc gctccggcgc gccgtgactg tcacctgcga agggattata atgaaacata 28260
acattgggtg gcttctaccc gccgccctcg cgacgcttgc cttcgtcccg gcctgcagcc 28320
cgaatcacgg tgaggatgcg ccctccgtga cgtcagcaga gagcggcgcg gcgccgagcg 28380
ctgactgcgt cgcgctcggg gcgaagctcc aggcggcgct ggacggcgcc gccgccgcgc 28440
aaaaggctcc gggagccgca gcggcggtcc agagcgggga ctgtgtctgg cggggcgcca 28500
cgggcgtctc ggacctggtc gcgagcacgc cgacgaagcc tggagatctc tttcggatcg 28560
gcagcatcac caagaccttc gtctctacgc tgatactcat gctccgggca gaaggccggt 28620
tgtcgctcga cgacgcggtg tcgaagtatg tgaagggcat ccccgccggc gaccagatga 28680
cgctgcgcca gatcctcggt cacacgagcg ggctcttcga ttacacgtac agcccggcgc 28740
tcggccaaat gatcgaggtg gatccgaccc gcgccttcgc gccggcagag ctcatcgccc 28800
tcgccacggc cgaggcgccg tatttcgcgc cgggcgcggg ttttcgctat tcgaacacca 28860
attacatcgt ggccggcctg gtggccgagg cggtgtcggg cgggacgctc gccgggctgc 28920
tccgcacgcg catcctagac cctgtgggcc tcgcgcacac gtatctggac ggcgccgagc 28980
cgccggtcca agggctcatc cgcggctacg gcgactacgg cgcgggcttg gtcgacatca 29040
ccgaccagct gtcgcccacc gaggcgtggg ccgccggcgc cctggtgtcg aacgtcgatg 29100
acctcaatcg cttctttgcc ctgctcatca gccacgagct gctctcgtcg gacgagcttc 29160
aggacatgac cacctggacc ccgacgatgt ggccccacga gcccggatat ggcctcggcc 29220
tcatcgagcg cgattctgcg ctcggctccc tcaacgggca ctgcggaatc atctgggget 29280
ttcaatcggc gtcgtacggg gtgcccggcc gcggcgacgc gatcaccgcg ctcatcaacc 29340
ggagcgacgg cgacgcagcg cggctcgtcg acgagctcgc gaaggtcgtg aaagagcgct 29400
gatcgaggcg gaatgggagc gcttcggcgg gtggtgatgg cgcccggcgc tcagaacgcg 29460
acgcgcagcc ccgcgctcag cgggcctgcg ccgggcgacg cggccacggc gcccggaccg 29520
acgaggagcc gcgcgacggc gggcgcgctc ggcgcgtcgt ctcgccgcac ccgccgcttg 29580
ccgaacacgt agagcggcag gccgacggcg accccggcca ccccgccgag cgcggtggcg 29640
atcgccacct cggacgcctc ggcgcgcgcg gcgctgctgt cgtcgtggct cgcgaagacc 29700
agcaccgcgc cgctgaggat ggcggcgccg cccagggtcg tgaggacgag ccccgagatc 29760
accatgaccg ggctgttcca ctccgtcgtc cgctcctcga agtcgcggaa cgccgcctcc 29820
gccgccgcga gctccagctc gatccggcgc tgctcggcgc gccgctcgtc cgcggagccg 29880
atctcgtgga cgcggcgggc ctggccctcc agcgccgcga tgcgcgcctc gtgcgcggcg 29940
gcggtctcct cccacgtggc ccctggcggg accccgccca cggccggcgc gacagagggc 30000
gccgacgccg gggtcgaggc gggcgccgcg ggcggctccg cggccacgga aggcgccgcc 30060
gccgcgggag gcgcgggcgg ctccgcggcc acggaaggcg ccgccgccgc ggaaggcgcg 30120
ggcgactccg cggccacgga aggcgcggcc gccgcgggag gcgcgggcgg ctccgctgca 30180
acggagagcg cgggcgccgc ggcagccagg cccagcgccc acgcgacgac acgacggcgc 30240
gccgcaaccg cgcgcgggcg cgcgaagcgg aggtggacct gctccatgcg cgcagcgtcg 30300
cccctcgaca gggccgggtc aaggcgcggg agtccgagag cacgagacct ccgcgccgca 30360

ggaaacaggc gcgccggcgg cccgcgcggc ggctcgccgc tcaccctcg cgcggccggc 30420
cgcggcgccg cctccctcc ccggcgggcc gcgcgtcggc ggccacgcgg agcagctcct 30480
ggaagtgccg ctccaccggg ccgaggtcga tgccgtccat gaacgacgtg aacgcgaagt 30540
acggcagcag cgtctgccag gcgcgcagcc agcccgggag gtagcgcggc cgctccagca 30600
cgccggcggc gcgcagctcc gccagcaccg cgatgtcgtc gagcgcgatc cggccgcaca 30660
cgagcagctc cacctcgtcg cgcagcccgc cgcgaaggac cgcggccagg aaggcgtaga 30720
cctcgagcag cccggccagg tagcaggcgt ccttggtgaa cggcgcgccg ccctcgacga 30780
gcccgccgcg gcacacgcgc tgggcgtcga agtaggcgtc gcggcgctcg gcgccgcgct 30840
cgcgcaggtg ccggtacagg tcgaggaagc tcgcgccctg ctcggccatg tccacgagcc 30900
gcacccgctc ggcgagccgg gtgaggcggc cgatggagag cgagcggctg tagagctcgg 30960
cgaagatggc gagcccctcc tgcgtgcgcg tggtgcgggg gccgcccgag cgcaggaacg 31020
cgcaccgcgg ctgcgccgcg ccgttgtgcg cggtgagcgc gtgcgtctcg acctcgtggt 31080
gccacagccc ctcggcctcc cacgccgcga aggtcgcctc cggccggatg cggacccggc 31140
tcatgccggc gaccaccttg gccgtgacgc gcgggtcgac ggtgatctcg aggtcgagcc 31200
gcggcgcccg gccggccacg cgggcggcga gcatgtcccg gagcgcgccg gcgtcgagcg 31260
gctcctcctc gggatcgctg gcctcgtccc agccgtggac gcgcaggcgc tcggtgaggt 31320
gctcggcgag gtcgatgttc ctgagcgagc cgccgaagaa ccgcgagcgc gcgccgccgt 31380
agagctcctg cgaccgcgcg gagaacgcgc gggtgcccgc ggcctcgagc agctccgcgg 31440
cctggatctg cgcgcggacg ttgtcccgca gccagccgag cgccggcgcg tccccgtcga 31500
tggccccgag gagctcgcgc agctcggcga cgcgccgcgc gaggccgtcg cgatcgacgc 31560
ggtactcgac ctcggggagg cggtcctcgc cggcggcgaa gaagcgctcc tccacctcgc 31620
gcggccaggc gatgtcctcg agcagcttga gggccttgcc ctccgccagg cggccgccca 31680
cccgatcgag ctgctccagc acggcgcggt cgatgctcat cgagcgcagg atcgccgaaa 31740
ccgcgagacg ccggaaccgt cattcctcg acgaggcagc gattgccatg ttccgtcgct 31800
ttttggagcg ccgtcgtcgc gctcgcctgc gggctccggc gatccagcgc ggttgcatgc 31860
agcgagggtg ttccggggct ggctcgagag cgtcctttgg cccacacccg agacacgaat 31920
gctccgcgcc gagcgcggtt gaccgtggac ccgccggaga gccgatgata cggtccggcc 31980
gatgtcggag agtgtagctc aactcgaaga acaccgcgcg gcgctcaccg gacactgcta 32040
ccggatgctg ggttcggtgg tcgacgccga cgacgccgtc caggagacga tggtgcgcgc 32100
ctggcggagc ctggataagt tcgacgggcg ctcgtcgctg cgcacctggc tgtaccgcat 32160
cgcgacgaac gtctgcatcg acctgcgggc cgaccgcgcg cgccgggcgc gccccatcga 32220
ggaaggcccg gtcggcacgg tggacgacgc gctcgagacg cgcccgcgca cccactggct 32280
ggagcccgtc cccgacgcgc acgccctgcc ggcggacatc gacgccgcgg agcgggcgat 32340
gctccgccag agcatccgcc tcgcgttcgt cgcggcgctc cagcacctgc cgccgaagca 32400
gcgcgccgcg ctgctgctca cggaggtgct cggctggtcc gccgcggagg tcgccgacag 32460
cctcaacacc tcggtcgccg cgatcaacag cgcgctccag cgcgcgcggg cgacgctcgc 32520
gagccgcgat ctcggcgacg cgcgcccctc gctgccggag ccgcagtccg cgctgctcga 32580
ccgctacgtc aacgccttcg agcggtacga cgtcgacgcg ctcacggcgc tgctgcacca 32640
ggacgcgacc ctgtcgatgc cgccgttcac cctgtggctc cgcggccacg agtcgatccg 32700
cgcctggctc gtgggcccgg gagcgggctg ccgcgggtcg cggctcatcc cgacggcggc 32760
gagcggctcg cccgcgttcg cgcagtatcg cccggcgccg gagggcggcc accgggcctg 32820
ggcgctcatc gtcctcgacg tcgcgggggа ccgcatcgtc agcatgacgt ccttcctcga 32880
caccgagacg ctcttcccgc ggttcggcct gccgctcgat ctaccggcgt agccgcgggc 32940
gccctgcctg cctcgccgcg ggtgccctgc ctgcctagcc gcgggcgccc ggcctggcca 33000
cgggcgcccg gcctggccac gggcgcccgg ccagcgacgg ggcgacgatt tttttctgag 33060
cgaccgatga gtcctgacgg ggccgggggt ctacgggggt gaatccaaca cggaggcacc 33120
catgaccgtg accatcgcca gcatcgatca tcgtgaccag gacctcatga ccgggccccа 33180
ggccaaggcg ccggcccgcg cggcggcgcc cgacgcggcg ccgtccaggc gagccgtgtg 33240

ggcgggccgc gtcctgagcg ggctggccac gctgttcctg acgttcgacg ccgcggtgaa 33300
ggtgctgaag ctgttccccg cggaggcgtc gaccgccgag ctcgggttcc cggcgcacct 33360
cgtccccacc ctcggctacc tccagatcgc ttgcctcgtg gcctacctga tcccgcgcac 33420
cgcggtgctc ggcgcgatcc tgtggaccgg ctacctgggc ggcgcgatcg cgatccacgt 33480
gcgggtcgag aacccgctct tcagccacac gctcttcccc atctacgtcg ccgcgttcct 33540
ctgggcgggg ctctggctgc gcgaccgccg cgtgcgcgcg ctgaccgcga gcccgtcgtc 33600
gcagggccga tgagcttcac gtttcacgag agtccatcac ggtaaaagga gaagcgagcc 33660
atgaccacaa agaacccccg caagctcttc gtcaacctgt ccgtccgcga cctgaagcga 33720
tcgatggagt tcttcagcaa gctcgggttc gagttcaacc cgcagttcac ggacgagaag 33780
gccgcctgca tggtcgtcag cgaggaggcc tatgtcatgc tcctcgtgga gtcgttcttc 33840
aagacgttca tgaagaagga gatctgcagc acgagcacgc acacggaagg gctcttcgcg 33900
ctctcgtgca gcagccgggc cgaggtcgac gacatggtga agaaggcggt cgcggcgggc 33960
gggtcgcacg cgatggatcc gcaggatcac ggcttcatgt acgggtggag cttctacgac 34020
gtggatggcc accactggga ggtcatgtgg atggatccca aggcgatcca gccgtagccg 34080
acggggctgg gcgcgccgcc tggaagagcc cccgtgaggc ggggaggcgg gaggatcacc 34140
gtcttcgtag cccacagcga tgcagtatcc gtcgcgcttc gtatcgaagc acggctgtta 34200
cgggcgcgtc agagcgcgtc gcaggtgatg ccgagccgca gcagcgacac gggcacgagc 34260
gtggctccga tggagatgag ccgagtctcg cccatggtct cggggtcatg aatggatgag 34320
taggggactc gctccttcgt cacgtcgtgc tcgacggcga cggcgaggcc gagctcgaag 34380
tgcacggggc ctggaccgaa gatccagctc gccccggcgc gagccccgac gaaaagcgtg 34440
tcgccgtcga cgccagggcc gtcgtcccag ccgggcgatc ccaccgcggt gtaggtgtgt 34500
ttcccgaagg aacccgcgag cgagagtcga agtccgaccg gcgctcgcca cgcgacgccc 34560
gctgtcgcgc cgacgccgcc gaagctctcc ccgaaaggct tatccctgt ctcgatgaag 34620
ccacccacct cgatgacgct gatgcggtac gtgagcgcga gattgaggtg cacccagcg 34680
ctgtccgagc ccgagtagag gccggcgccc acctgcacgc tgaaatccat gctcggcgcg 34740
gatccgcgcg caggagcgac gccaggggcg ctgccctcct gcgcgcgggc cgtcccgacg 34800
caaagaaaga gggctgtcgc gaagaatcca agcgagatcg atcgaagtga gcgcatgtcg 34860
ggccctggag catccgctgt accaggtgcg tcgtattcat gcggcgcgcc gccgggcgcc 34920
gccgcgctgg cctgtccgac gcgagatcac gaatccgcca tcgctcccct gggccgccgg 34980
ccgctctggt tcgcctgcgg gcgtgcgccg gcgctcgtgt ggcccatggc aaccttgtcg 35040
cggtgtcgct cgaacagcac agagagtatc gcgtccgcaa caaccgcgcg acccggcgag 35100
acgctcgtgg ggccccctgc ctccccactt catcataacg ccatcaggag cactcgacat 35160
ttcatttctt cacctccact ggctgagggc gacggtgctc gtcatcggcc ggttgctctg 35220
gcggttgctc tggcggggtt tctgacgccc ggaactaacg cttcgagcgc tcccccttgc 35280
tctcccgttc cttcagctcc tccagcaggt cgtcgaggcg ctcgtagctg cctttccaga 35340
agcggcggta gttgtcgagc cagccgctgg cgtcctcgag cggcttggcc tcgatccgac 35400
aaggcctccg ctgcgcgtcg cggccgcgcg agatcaggcc cgctcgctcc agcaccttga 35460
ggtgcttgga gatcgcgggc tggctcatcg cgaacggctt cgccagctcg gtcaccgacg 35520
cctcgccgga cgcgaggcgc gcgaggatcg ctcgccgtgt cggatcggcg agcgcagcga 35580
acgttgcgtc gaggcgctcg gacggggtca ttgcataact ccttggtata aaaaccagtt 35640
agttatacaa cctggggccc gggcggtcaa gcctccaggc gatggcggtt cggcccgggg 35700
gctccgctcg cggcacgcgc gccgcgcggc tacgtgcgcg gcgcggtgag cacgtcctgc 35760
agcgtggcgc cgaccacggg cttggtcagg tgcaggtcga agccggcccg cctggacctg 35820
gcctgatcgt cgggcccgcc gtagcccgag agcgccacca ggtagagcgc ttcgccgccg 35880
ggcgcggccc gcgcccggcg cgcgacctca taaccgtcga tgccgggcaa gccgatgtcc 35940
acgaaggcca cctcgggggcg cagctccaga agcttcttca cgcgctccag cccgtccacc 36000
gccaccgtca cctcgtgccc cagcgcctcg atgtacgccc gcatcacccg gcgcacgtcc 36060
tccgcgtcct ccacgacgag cacccggcgc cggtcagccg ccgcctcggg cgcctcggcg 36120

```
cgctgcgccg gaggcggcgg cggctcgtcg cgctgcgccg gaggcggccc ctcgcgcggc 36180
gggggcggcc cggcgctcgg ggcaggctgc ggcgccgccc cggggccgag cggcaggcgc 36240
acggtgaact cgctgccctg gcccggcccg gcgctcgccg cggccacgct gccgccgtgc 36300
agttccagga gccgccgcac cagcgtgagc ccgagcccca gcccgcccgt gctccggtcg 36360
atggtctggt cgaccctgcgt gaacagatcg aacaccttct cgagcatcgc cgccgggatg 36420
ccgcggcccg tgtcgcgcac ccgcagcacg gcctcgggcg cgccgaccgc cgcctcgcgc 36480
gtgaggcgca ccgagatcga gccccccggc ggggtgtact tcgcggcgtt ggtcaggagg 36540
ttcgtcacca cctgctccag ccgcgtcgcg tcggcccgca tgccgaagtc cccgggcccc 36600
accgacagcg acacgtcatg gcgccgggcc tcgacggccg gcctcaccgc ggcggcggcg 36660
ctctgcacca ccgccgcgag atcgacgtcc tcgaggcgca gctccaccgt gccccgcgtg 36720
atgcgcgaca cgtcgagcag atcgtcgacc agccgcacga ggtggcccat ctgccgccgc 36780
gcgatctccc ggtagcgcgc cgacgcgggc ccgtcgccgt ccgcgtcgtc gagcagcgtc 36840
agcgacaggc tgatcgaggc catcgggttc cggagctcgt gcgcgagcat cgcgaggaac 36900
tcgtccttgc gctgatcggc gagcttcagc gcctcgacga gcgcctccac gcgcctccgg 36960
gcgcgcacct ggtcggtcac gtcgaacgcg aacacgaaga cgccctcgac cgccccgtcg 37020
cgatcgcgca tcggctggta gacgaagttg aagaacacct cctccgtcgt gccgtcgccc 37080
cggcgatcga gccgcaccgg gagctccttg ccgacgatgg gctcgccggt gcggaccacc 37140
gcgtcgagga gctcccagat gccctgtccc tcgagctcgg ggagggcggc ccggatgggc 37200
tcgcccacga gcgatcgacc gccgacgagc cgctggtaga gcgggttgac cacctcgaag 37260
acgtgctccg gcccgcggag gatggcgatg ggccccgggg cctgcatgaa gaggtcgttc 37320
aggtactggc gctgcccctc ggcctcgcgc cggcggcgcg cgagctcgac gtggatgcgg 37380
acccgcgcga ggagctcctt cgcggagaac ggcttcacga ggaagtcgtc ggcgccggcc 37440
tcgaggctgt cgacgcgcgc ctcctcgccc gcgcgcgcgg agagcatcac cacggcgacg 37500
ccgcgggtgc gatcgtcggc gcgcagcgcc ctgagcaggc cgaagccgtc gagccgcggc 37560
atcatcacgt cggtgagcac gagatccggc gggtgggcgc gggcgcgctc cagggcggcc 37620
cgaccgtcgg ccacggcctc caccgtccac ccctccgcca cgagcagccg cagcgcgtac 37680
tcgcgcatgt ccgcgttgtc gtcggcgacg aggacgcgcc ccggcagcct cccggccggc 37740
ccctcgcccg ccggccggga cgccggcgcc tgctcgccgc ggagccactg cgcggcctcg 37800
tcgaggaagg gcgcggcgtc ccgccccccc gcggccggcg ccgaggccgg cgcgac     37856
```

Patentansprüche

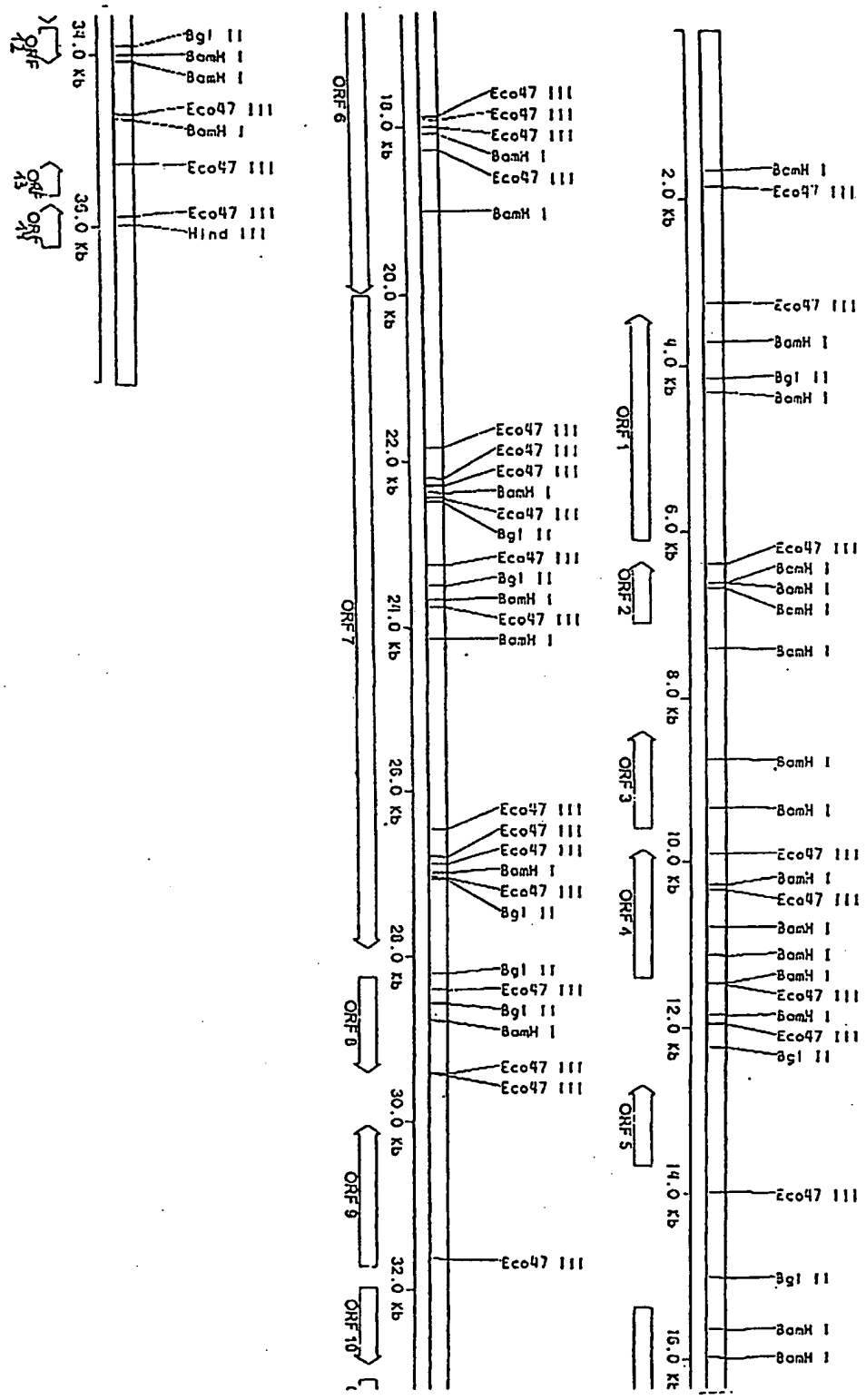
1. DNA-Sequenz, deren Expressionsprodukte die enzymatische Biosynthese, Mutasynthese oder Partialsynthese von Polyketid- oder Heteropolyketidverbindungen bewirken oder daran beteiligt sind.
2. DNA-Sequenz nach Anspruch 1, wobei es sich bei den Polyketid- oder Heteropolyketidverbindungen um Epothilone handelt.
3. DNA-Sequenz nach einem der vorhergehenden Ansprüche, wobei die DNA-Sequenz Regulationselemente (ORF9, ORF11 und ORF12) und Transkriptionsregulatoren (ORF10, ORF13 und ORF14) aufweist und die Expressionsprodute eine tRNA-Synthetase (ORF1), Monooxygenase (ORF2), Aminotransferase (ORF3), Tyrosin/DOPA-Decarboxylase (ORF4), 3-Oxoacyl-ACP-Reduktase (ORF5), Polyketidsynthase (ORF6), Peptidsynthetase (ORF7) und Transpeptidase (ORF8) umfassen.
4. DNA-Sequenz nach einem der vorhergehenden Ansprliche, wobei die DNA aus Myxobakterien stammt.
5. DNA-Sequenz nach einem der vorhergehenden Ansprüche, wobei die DNA aus Sorangium-Stämmen stammt.
6. DNA-Sequenz nach einem der vorhergehenden Ansprliche, wobei die DNA aus Sorangium cellulosum stammt.
7. DNA-Sequenz nach einem der vorhergehenden Ansprüche, wobei die DNA ausgewählt ist unter:
   (a) der folgenden DNA-Sequenz: oder deren komplementärem Strang,
   (b) DNA-Sequenzen, die unter stringenten Bedingungen an die Proteine kodierenden Regionen der in (a) definierten DNA-Sequenzen oder an Fragmente davon hybridisieren,
   (c) DNA - Sequenzen, die wegen der Degeneration des genetischen Kodes an die unter (a) und (b) definierten DNA-Sequenzen hybridisieren,
   (d) alle Variationen und durch Substitution, Insertion oder Deletion von Nucleotiden entstandene Mutanten der unter (a) bis (c) definierten DNA-Sequenzen, die isofunktionelle Expressionsprodukte ergeben.
8. Rekombinierter Expressionsvektor, der eine DNA-Sequenz nach einem der Ansprüche 1–7 enthält.
9. Prokaryotische oder eukaryotische Zelle, die mit einer DNA-Sequenz nach einem der Ansprüche 1 bis 7 oder mit einem rekombinierten Expressionsvektor nach Anspruch 8 transformiert oder transfiziert ist.
10. Zelle nach Anspruch 9, wobei die Zelle von Myxobakterien stammt.

11. Zelle nach Anspruch 9, wobei die Zelle von einem Sorangium-Stamm stammt.
12. Zelle nach Anspruch 9, wobei die Zelle von Sorangium cellulosum stammt.
13. Zelle nach Anspruch 9, wobei die Zelle von einem Streptomyces-Stamm stammt.
14. Zelle nach Anspruch 9, wobei die Zelle von Escherichia coli stammt.
15. Verfahren zur enzymatischen Biosynthese, Mutasynthese oder Partialsynthese von Polyketid- oder Heteropolyketidverbindungen, bei dem eine Zelle nach einem der Ansprüche 9 bis 14 in einem geeigneten Kulturmedium kultiviert und die Polyketid- oder Heteropolyketidverbindung aus dem Medium isoliert wird.
16. Verfahren nach Anspruch 15, wobei die Polyketid- oder Heteropolyketidverbindung ein Epothilon ist.

Hierzu 2 Seite(n) Zeichnungen

Nummer: DE 198 46 493 A1
Int. Cl.[7]: C 12 N 15/52
Offenlegungstag: 13. April 2000

Fig. 1



Nummer: DE 198 46 493 A1
Int. Cl.[7]: C 12 N 15/52
Offenlegungstag: 13. April 2000

Fig. 2

	Gen/Funktion	Position
ORF 1	tRNA-Synthetase	6100-3397
ORF 2	Monooxygenase	7111-6374
ORF 3	Aminotransferase	9550-8433
ORF 4	Tyrosin/DOPA-Decarboxylase	11393-9854
ORF 5	3-Oxoacyl-ACP-Reduktase	12212-13656
ORF 6	Polyketidsynthase	15374-19984
ORF 7	Peptidsynthetase	20003-27889
ORF 8	Transpeptidase	28251-29402
ORF 9	Regulationselement	31720-30040
ORF 10	Transkriptionsregulator	31982-32932
ORF 11	Regulationselement	33128-33613
ORF 12	Regulationselement	33661-34076
ORF 13	Transkriptionsregulator	35616-35255
ORF 14	Transkriptionsregulator	36242-35730